

Establishing AI Accountability in Policing and Security

BAYERL, Petra <<http://orcid.org/0000-0001-6113-9688>> and AKHGAR, Babak <<http://orcid.org/0000-0003-3684-6481>>

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

<https://shura.shu.ac.uk/38089/>

This document is the Published Version [VoR]

Citation:

BAYERL, Petra and AKHGAR, Babak (2026). Establishing AI Accountability in Policing and Security. In: AKHGAR, Babak and BAYERL, Petra, (eds.) Accountability for Artificial Intelligence in Policing and Security. Security Informatics and Law Enforcement . Cham, Springer, 71-84. [Book Section]

Copyright and re-use policy

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



Establishing AI Accountability in Policing and Security

Petra Saskia Bayerl and Babak Akhgar

5.1 INTRODUCTION

Artificial Intelligence (AI) presents policing with important capabilities to increase the speed, accuracy, and range of investigations whilst reducing costs and potentially directing resources more effectively and efficiently. Some AI models provide the potential to solve old crimes by helping investigators to identify cases that contain promising evidence which could be re-examined using new forensic techniques. They also help to prevent new crimes, particularly where technology is being used by the perpetrators, for example, to detect cyber-attacks faster, identify large-scale frauds, or fend-off disinformation campaigns, or detect deep fake. In fact, AI use becomes an imperative given that criminals and crime networks increasingly deploy AI against society and States (Chaps. 13, 14).

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

© The Author(s) 2027
B. Akhgar and P. S. Bayerl (eds.), *Accountability for Artificial Intelligence in Policing and Security*, Security Informatics and Law Enforcement, https://doi.org/10.1007/978-3-032-41284-3_5

At the same time, there are understandable concerns in society about the police use of AI that are centred around the infringements of privacy, fundamental freedoms, or even critical thinking (e.g. [8, 12]). These concerns also find expression in an increasing tendency to challenge AI use in the courts and the tightening of regulations on national and international levels. The challenge for police is how to capitalize on AI to safeguard society whilst also assuaging societal concern about their use of advanced technological capabilities.

We argue that AI accountability is an important concept that binds organisations to enforceable obligations and thus ensures actionable procedures to assess, evidence, and enhance responsible AI usage in policing. Beyond the concepts of trustworthiness and explainability, AI accountability imposes obligations on organisations to justify their AI usage, monitor its application, and above all, enforce changes and redress if things go wrong.

Whilst accountability seems an abstract concept, it can be a very practical instrument, not only for the proactive risk assessment and management of AI but also to understand the benefits of AI deployment: *‘It helps to think through the why, what, how, who and for whom of AI deployments, and it does so proactively. It supports police forces to: understand risks and strengths before each AI procurement, deployment or change; know up-front what to do if something goes wrong; [and] understand which areas of accountability are achieved or need addressing, and have evidence ready in case they get challenged and have to proof adequate AI procedures’* [5, p. 17].

This chapter offers firstly the conceptual foundation for AI accountability in policing, in that it defines AI accountability as a composite of 12 accountability principles. It further explains how to make the concept practical and assessable through the operationalisation of the principles into eight themes.

5.2 DEFINING AI ACCOUNTABILITY IN A POLICING CONTEXT

Accountability is the ‘acknowledgement and assumption of responsibility for actions, decisions, and their consequences’ [10] and comprises in itself the three aspects of monitoring, justification and enforcement [15]. Accountability thus offers the ability to monitor and enforce obligations towards one or multiple stakeholders, in the understanding that not meeting these obligations will lead to consequences.

The importance of accountability lays in its aim of ‘preventing and redressing abuses of power’ [15, p. 14]. It has thus at its very core the notion of negotiations across various legitimate interests when it comes to the deployment of AI for policing and security purposes.

AI Accountability specifically refers to ‘the acknowledgement of an organisation’s responsibility, when using AI, to act in accordance with the legitimate expectations of stakeholders and the acceptance of the consequences—legal or otherwise—if they fail to do so’ [2].

5.2.1 *The Need to Consider the Complete AI Lifecycle*

Often, AI accountability discussions are focused on algorithms (e.g. [9]), the purchase of AI capabilities, or deployment stages [1]. However, accountability does not stop once a system is no longer in use. Whilst the facial recognition systems used by the Metropolitan police to identify persons wanted for previous serious criminal offences or individuals representing a risk of violence at public events may only be in use for some hours or days,¹ consequences may last considerably longer—both in a positive sense such as arrests of wanted offenders and in a negative sense—when decisions may have been wrong infringing on the rights of the individual affected. The Harm Assessment Risk Tool (HART) used by Durham Police in the UK to assess the likelihood of suspects for reoffending was heavily criticised for being flawed and to systematically over-estimate the risk of re-offending using ‘racist data profiles’ [11; also 7].

AI capabilities leave data, decisions, and impacts, even if they are decommissioned. To ensure a comprehensive risk assessment and management, a core requirement is therefore that police organisations evaluate AI accountability across the full lifecycle of their AI capabilities or systems. The lifecycle begins before deployment, with either design or procurement, and ends not with deployment but with post-deployment to handle the legacy of AI capabilities (see Fig. 5.1). We hence argue that the full lifecycle should be conceptualised in six stages:

1. **Pre-deployment:** Encapsulates all considerations and decisions before an AI system or capability enters operational use. This

¹ <https://www.met.police.uk/police-forces/metropolitan-police/areas/about-us/about-the-met/facial-recognition-technology/>.

encapsulates disparate ‘origins’ of the AI system, as either self- or co-developed by a police force or as procured from ‘outside’, creating two entry points:

- a. *Research and design*: the AI capability is created or co-created with direct input from the adopting police force; this includes full consideration for training datasets
 - b. *Procurement*: the AI capability is brought in (e.g. purchased) from a third-party
2. **Deployment**: Encapsulates all considerations and decisions during the operational deployment of an AI capability
- a. *Operational usage*: running of the AI capability according to a specified purpose
 - b. *Revision*: adaptation of the AI capability either in (a) its features or (b) in the purpose of its application
3. **Post-deployment**: Encapsulates all considerations and decisions after the operational deployment of the AI capability has ended
- a. *Migration*: the AI capability is replaced by a different AI capability for the same or closely related purpose and data, decisions, etc. are transferred
 - b. *Decommissioning*: the AI capability is retired, and no further AI capability is used going forward

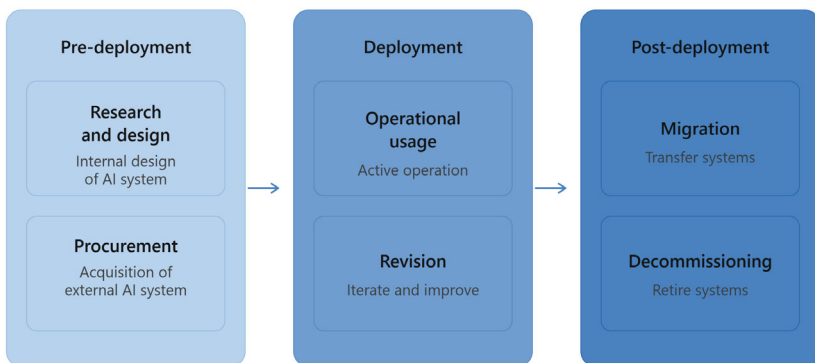


Fig. 5.1 AI lifecycle stages

This comprehensive view of the AI lifecycle is vital, as each stage is linked to disparate AI accountability considerations, roles and responsibilities. For instance, a police force (co)creating an AI system has considerable influence on the exact design but requires strong internal expertise on algorithms, models, and training data. In contrast, the procurement of a system means that numerous decisions may not be in the hand of the deploying organisations (e.g. training data or algorithms used), which shifts responsibilities for ensuring and evidencing AI Accountability to provider organisations (Chap. 7). Similarly, deployment and post-deployment are likely to involve disparate core functions such as analysts during deployment, and data protection officers during post-deployment.

Since each stage in the AI lifecycle is linked with specific considerations and decision points, they each require stage-specific planning and management (for details see the AIPAS Implementation Guidance in the Appendix). Particularly the transition phases—from pre-deployment to deployment and from deployment to post-deployment—require careful management, when responsibilities, knowledge, data, logs, etc. may need to be handed over to new functions.

5.2.2 *AI Accountability as a Process and Network of Mutual Obligations*

The AI lifecycle, whilst important, is still a largely static perspective on accountability, where AI accountability responsibilities and tasks transition (more or less) neatly from one stage to another. Whilst true in the broader sense, in practice accountability is much ‘messier’. Our concept of AI accountability is therefore based on two core understandings: that *accountability is a process* and that *it is accomplished in a network of mutual obligations* [5].

5.2.2.1 *Accountability as a Process*

AI deployments require active management to ensure that they remain legal, transparent, responsive to, and in line, with societal expectations. AI accountability is not something that can be assessed once and then assumed. It needs to be assessed and re-assessed across the lifecycle of an AI system. Also, external triggers can require a review: new policies or regulations may change the legal landscape for data collection or

retention, new AI capabilities may require a review of decision oversight, etc.

Moreover, accountability obligations such as learning and redress (see Sect. 5.3.1) are by design intended to ensure that accountability arrangements, processes, and mechanisms can adapt to changing circumstances. Transparency, explainability, and inclusivity are further accountability requirements that emphasise process, as communication and negotiation with diverse stakeholder groups is seldom straightforward but needs continuous engagements and refinements.

5.2.2.2 *Accountability as a Network of Mutual Obligations*

Policing accountability implicates a multitude of functions within the police, as well as external stakeholders the police are accountable to (e.g. ministries, mayors, general public) and accountable for (e.g. victims of crime, communities at risk) (Chaps. 2, 3). The introduction of AI does not change this but rather adds additional obligations, such as avoidance of algorithmic biases, automation biases, and human oversight for decisions.

However, we argue that accountability is not only from police to others; instead, there are also obligations that other stakeholders have towards the police. An example are the responsibilities of AI providers to be transparent about training data and faults in algorithms of their systems that police may purchase. Another example is the reliance of policing on the public for support such as the reporting of crimes, information relevant to investigations, or even simply the adherence to laws. As we formulated elsewhere, *‘the creation of such relationships may well carry a legitimate expectation on behalf of internal security actors that citizens will attract some degree of accountability for the data they contribute’* [4, p. 4].

Mapping out a comprehensive view of relevant stakeholders for a specific (planned or current) AI deployment is crucial to understanding these mutual obligations. We refer to this comprehensive overview as the ‘stakeholder landscape’ that allows for the systematic outline of which functions or organisations are accountable to whom, in which way, and what respective requirements and needs are for information, communication, engagement, evidence, etc. This enables the police to target accountability measures in the correct way to ensure all stakeholders are able to understand the purpose, remit, and (potential) impacts of

AI deployments. The objective is to make the AI accountability process practical, so it can be translated into concrete steps and measures.

5.3 MAKING AI ACCOUNTABILITY PRACTICAL AND ASSESSABLE

AI accountability is not separate from police accountability; if done right, it is an integral part of police accountability and every-day accountability arrangements.

Police accountability is a broad and complex concept that does not lend itself easily to assessment and quantification. A common approach to make police accountability more accessible is a principles-based approach. Principles have been used in numerous AI ethics frameworks, such as OSCE or ALTAI (Assessment List for Trustworthy AI [13]), where abstract concepts such as responsible AI or trustworthy AI are broken down into their defining components (i.e. principles). ALTAI, for example, decomposes trustworthy AI into seven principles ranging from *Transparency* and *Diversity*, *non-discrimination* and *fairness* to *Human Agency and Oversight*, with each principle further specified by questions that aim to assess whether this principle has been fulfilled.

The same approach has also been fruitfully applied to policing accountability in the UK. The framework by Lennon et al. [14] who proposed 12 principles that together create a conceptual basis to establish police accountability.

Using an earlier version of this framework as a foundation, we engaged in a bottom-up process that contextualised police accountability to AI in a policing and security context. The AP4AI (Accountability Principles for AI [2]) project was framed within a European context with the aim to define AI accountability in a way that makes it translatable into practical mechanisms, as well as transferable across deployment and national contexts. The process involved extensive consultations with over 100 subject-matter experts from law enforcements, courts, ethics, industry, and civil society across 28 countries to identify the relevant ‘universal’ components (i.e. principles) of AI accountability in policing (details of the process can be found in [2, 3]). The result was a set of 12 AP4AI principles, which were further validated in an international citizen consultation with over 6,500 citizens in 30 countries (all 27 EU Member States, UK, US, Australia [6]).

This foundation was further developed in the subsequent UK-based project AIPAS (AI Accountability for Policing and Security)² in which the 12 principles were re-validated and refined for the national UK context, as well as four specific policing areas: forensics, counter terrorism, serious and organised crime, and child sexual exploitation (see Chap. 8 for details). Below we detail the final AI accountability principles and their relevance for the AI accountability process.

5.3.1 *The 12 AI Accountability (AIPAS) Principles*

The final set of AI accountability principles consists of 12 components that together allow to describe the full scope of obligations and requirements across the AI lifecycle and with explicit consideration of needs by all relevant stakeholders. The following list gives the short definition of the AIPAS principles:

1. **Lawfulness—‘follow the law’:** All aspects of the use of AI should be lawful. Lawfulness includes compliance with specific legal requirements such as the EU AI Act (where applicable), as well as organisational and professional policies which must be clearly identified and readily available
2. **Completeness—‘leave nothing out’:** AI accountability must cover all relevant aspects of the AI ecosystem and lifecycle including its design, supply, deployment, regulation, monitoring, and impacts
3. **Inclusivity—‘leave no one out’:** AI accountability arrangements must involve all relevant stakeholders engaged in and affected by an AI deployment
4. **Transparency—‘be open’:** AI accountability needs clear, accurate and meaningful information (subject to operational sensitivities) to enabling those directly and indirectly affected or responsible for oversight to make informed judgments and risk assessments about the use of an AI system and the accountability arrangements and support learning
5. **Impartiality—‘empower independence’:** AI accountability needs impartial and independent monitoring, judgement and redress

² AIPAS (‘AI Accountability for Policing and Security’) was funded by UKRI Engineering and Physical Sciences Research Council (Responsible AI (RAI) IA091 Grant Ref: EP/Y009800/1). For more details see: <https://centric-research.co.uk/projects/aipas>

mechanisms, without any conflict of interest in a personal, political, financial and functional way. Any conflict of interest must be identified and addressed

6. **Proof—‘follow the evidence’:** a forensic approach to all aspects of AI systems and of the accountability process itself, demanding and following clear evidence; The quality of that evidence should reflect the potential impact of the AI system’s use/non-use and mirror the standards of operational evidence gathering in terms of integrity, credibility and continuity
7. **Enforceability and Redress—‘make it right’:** external legal and procedural routes for complaint and challenge, but also internal mechanisms for individual enforceability and redress (such as professional and policy standards) and contractual arrangements
8. **Compellability—‘make it work’:** oversight bodies must be in position to make the AI accountability arrangements work giving relevant bodies the ability to compel the sharing of necessary information and evidence
9. **Explainability—‘describe, demonstrate, demystify’:** provide information about AI capabilities, their purpose, and impacts in a meaningful way that is easily understood by the relevant participants/audience
10. **Constructiveness—‘aim for better’:** stakeholders participate constructively with a shared aim of improvement; this may include considering different perspectives, inviting challenge and recognising how disagreement can lead to beneficial solutions
11. **Conduct—‘hold yourself accountable’:** relates to adherence to professional standards, values and expected behaviours which incorporate integrity and ethics
12. **Learning—‘look for the lesson’:** willingness of organisations and people to improve AI in every respect through the application of (new) knowledge and insights, including the modification and improvement of systems, structures, practices, processes, knowledge and resources, as well as the development of professional doctrine and agreed standards

The principles have connections to each other. For instance, *Explainability* is linked to *Inclusivity*, as the way *Explainability* is enacted depends on the number, type, and needs of the broad stakeholder

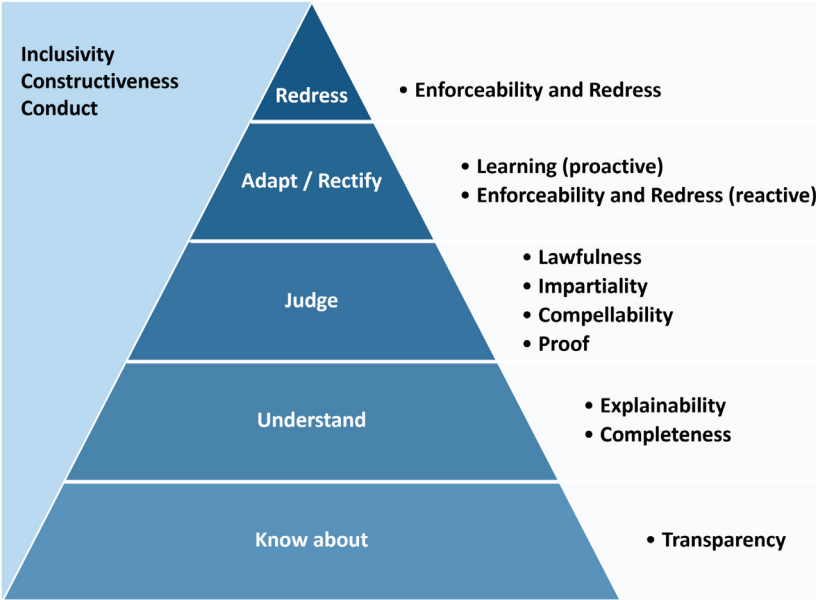


Fig. 5.2 Relevance of the principles for the AI accountability process

landscape. In a similar way, *Learning* as well as *Redress* require *Constructiveness*, if the police aim to meaningfully engage with lessons from problematic or faulty AI deployments and with individuals/groups that have encountered negative consequences. However, the 12 principles address distinct aspects in the accountability process (Fig. 5.2), which means all principles are required to fulfil policing’s AI accountability obligations.

5.3.2 *Operationalisation in eight themes*

To make AI accountability meaningful, it is necessary to assess how well its different principles and obligations are fulfilled. An important part of the AIPAS project was the operationalisation of the 12 principles into discrete criteria that can be judged and evaluated. We did so as a set of targeted questions that establish whether AI accountability is achieved in a specific area. The questions are formulated in a way that they can be answered as

‘yes’ or ‘no’, which allows the systematic, quantitative analysis of how far AI accountability achieved, as well as the identification of current gaps.

Across the 12 principles, the questions cover eight themes that cluster the principle-related questions according to the type of information required:

1. **AI system:** defines what the AI system does, how it functions, its purpose, limitations and who is responsible for its development and oversight
2. **Data:** focuses on the types, sources, and handling of data used in the design and deployment of the AI system
3. **Laws and regulations:** ensures that the use of AI complies with applicable laws, codes of practice, regulatory requirements and policies
4. **Risk assessment and management:** identifies risks, harms or biases associated with the AI system and implements mitigation measures
5. **Oversight and redress process:** establishes mechanisms for independent scrutiny and AI accountability, allowing stakeholders, particularly the public, to challenge decisions and seek corrections, ensuring just outcomes
6. **Stakeholders:** involves mapping and engaging all relevant parties who are involved in or affected by the AI system
7. **Accountability evidence:** focuses on documenting decisions, processes and actions taken to prove AI accountability
8. **Awareness and learning:** encourages continuous improvement through feedback, review, and learning

The benefit of the theme-based approach for end-users is that it collects AI accountability information in a more familiar ‘policing’ way. The translation of principles into themes and concrete questions further offers a standardised approach to AI accountability assessments, which is a foundation for the implementation of software-aided assessments solutions as presented in Chap. 9.

5.4 CONCLUSION

This chapter offers a practitioner-focused approach to define and implement accountability for AI in policing. The AI accountability concept presented here has been co-developed with policing practitioners, as well as experts in relevant fields from law to ethics and industry, ensuring that it is comprehensive and relevant for strategic, tactical and operational levels of AI deployments. This approach has been validated in Europe and the UK, as well as with subject-matter experts and citizens across more than 28 countries. These findings show that it is ‘jurisdiction-neutral’, i.e. works across contexts and deployment scenarios.

Our approach AI accountability is specifically aimed at the practical implementation in policing. The subsequent chapters offer concrete advice on how to implement AI accountability across specific topics (organisational practices: Chap. 6, lifecycle, stakeholder and citizen considerations: Chaps. 7, 8, 10), as well as software-supported solutions (Chap. 9).

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. A. Agarwal, H. Agarwal, A seven-layer model with checklists for standardising fairness assessment throughout the AI lifecycle. *AI Ethics* **4**, 299–314 (2024). <https://doi.org/10.1007/s43681-023-00266-9>
2. B. Akhgar, P.S. Bayerl, K. Bailey, R. Dennis, S. Heyes, A. Lyle, A. Raven, F. Sampson, M. Gercke, Accountability principles for artificial intelligence (AP4AI) in the internal security domain: summary report on expert consultations (AP4AI Report 1) (CENTRIC, 2022). Retrieved from: https://ap4ai.eu/sites/default/files/2023-09/AP4AI_Report1_27Jan2022.pdf
3. B. Akhgar, P.S. Bayerl, K. Bailey, R. Dennis, H. Gibson, S. Heyes, A. Lyle, A. Raven, F. Sampson, Accountability principles for artificial intelligence (AP4AI) in the internal security domain: AP4AI framework blueprint (AP4AI Report 2) (Europol Innovation Lab, 2022). Retrieved from: https://ap4ai.eu/sites/default/files/2023-09/AP4AI_Framework_Blueprint_22Feb2022.pdf

4. B. Akhgar, P.S. Bayerl, G. Mounier, R. Linden, B. Waites, AP4AI: accountability principles for artificial intelligence in the internal security domain. *Eur. Law Enforcement Res. Bull.* **22**(6) (2022) <https://bulletin.cepol.europa.eu/index.php/bulletin/article/view/567>
5. B. Akhgar, P.S. Bayerl, How to do AI accountability—and why it’s worth the effort, in *The Police Foundation. Policing and the Fourth Industrial Revolution. Conference Report of the Cumberland Lodge Police Conference* (2024), pp. 16–17
6. P.S. Bayerl, M. Obst, B. Akhgar, International Citizen Consultation on AI Accountability in Policing (AP4AI Report 3) (CENTRIC, 2023). Retrieved from: https://ap4ai.eu/sites/default/files/2023-10/AP4AI_International-Citizen-Consultation_October2023%28online-version%29.pdf
7. B.B. Watch, A Closer Look at Experian Big Data and Artificial Intelligence in Durham Police. Online Article published (6 Apr 2018) (2018) <https://bigbrotherwatch.org.uk/blog/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/>
8. K. Blount, Using artificial intelligence to prevent crime: implications for due process and criminal justice. *AI Soc.* **39**, 359–368 (2024). <https://doi.org/10.1007/s00146-022-01513-z>
9. F. Cech, The agency of the forum: mechanisms for algorithmic accountability through the lens of agency. *J. Respons. Technol.* **7–8**(2021), 100015 (2021). <https://doi.org/10.1016/j.jrt.2021.100015>
10. R.A. Duff, Moral and Criminal Responsibility: Answering and Refusing to Answer (2017). <https://ssrn.com/abstract=3087771>
11. Fair Trial, FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool. Online Article published 15 Aug 2024 (2022). <https://www.fairtrials.org/articles/news/foi-reveals-over-12000-people-profiled-by-flawed-durham-police-predictive-ai-tool>
12. M. Gerlich, AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies* **15**(1), 6 (2025). <https://doi.org/10.3390/soc15010006>
13. High-Level Expert Group on Artificial Intelligence (AI HLEG), The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (2020). <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>
14. G. Lennon, N.R. Fyfe, J. McNeill, F. Sampson, *Principles for Accountable Policing* (The Police Foundation, 2023) https://www.police-foundation.org.uk/wp-content/uploads/2010/10/accountable_policing_FINAL.pdf
15. A. Schedler, Conceptualizing accountability, in *The Self-restraining State: Power and Accountability in New Democracies*, ed. by A. Schedler et al. (1999), pp. 13–28

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.

Babak Akhgar is Professor of Informatics, Director of CENTRIC and a Fellow of the British Computer Society. He has extensive and hands-on experience in the development, management and execution of KM projects and large international security initiatives (e.g. combatting terrorism and organised crime, AI in Policing, cyber security, public order and cross-cultural ideology polarisation). He has more than 100 refereed publications in international journals and conferences on information systems, specifically focusing on knowledge management (KM). Babak was awarded an OBE in June 2022, the second highest order of the British Empire award, in recognition of his services to security research.

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.

