frontiers in Political Science

TYPE Original Research
PUBLISHED 15 July 2026
DOI 10.3389/fpos.2026.1812399

OPEN ACCESS

EDITED BY
Leslie Paul Thiele,
University of Florida, United States

REVIEWED BY
Pitshou Moleka,
Eudoxia Education Private Limited, India
Swetha Sistla,
FinTech, United States
Mohd Arif,
Galgotias University, India

*CORRESPONDENCE
Jorge Constantino
✉ j.constantinotorres@gmail.com

RECEIVED 16 February 2026
REVISED 05 June 2026
ACCEPTED 08 June 2026
PUBLISHED 15 July 2026

CITATION
Constantino J, van Krimpen F, van Gerwen S and Wagner B (2026) Accountability perspectives in the context of AI in intelligence and security investigations.
Front. Polit. Sci. 8:1812399.
doi: 10.3389/fpos.2026.1812399

Accountability perspectives in the context of AI in intelligence and security investigations

Jorge Constantino[1]*, Floris van Krimpen[1], Sarah van Gerwen[2] and Ben Wagner[1,3,4]

[1]Technology, Policy and Management, Delft University of Technology, Delft, Netherlands, [2]Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [3]Applied Research Creativity Research Center, Linz, Austria, [4]InHolland, The Hague, Netherlands

This paper examines how democratic intelligence and security actors understand accountability in the context of AI-supported digital investigations. Against a backdrop of escalating AI-enabled cyber threats and growing governmental interest in adopting AI for defense, security, law enforcement, and justice, the potential use of AI raises pressing questions regarding accountability in democratic societies. Drawing on exploratory qualitative research based on semi-structured interviews with intelligence and security actors primarily in the Netherlands, triangulated with perspectives from Belgium, Norway and the United Kingdom, the findings show that accountability challenges are understood less as problems arising from AI alone and more as broader socio-legal-technical challenges involving AI and data governance, technological trustworthiness, human oversight, and organizational resources. Participants also emphasized that accountability in the digital age requires intelligence and security actors to navigate the persistent tension between protecting national security and respecting fundamental rights. The article provides empirical insight into how accountability is perceived and operationalized in AI-supported intelligence and security investigations under conditions of legal, organizational and technological complexity.

KEYWORDS
accountability, artificial intelligence, data governance, fundamental rights, generative AI, intelligence and security organisations

1 Introduction

Recent reports indicate that threat actors are utilizing Artificial Intelligence (AI) tools, including Generative AI (GenAI), to generate malware code or phishing email content. Threat actors can instruct these AI models to find sensitive user data and exfiltrate it.[1] The European Council's documents note the rising complexity of cyber threats as adversaries exploit advanced techniques that compromise citizens' privacy and critical infrastructure, affecting national security interests.[2]

1 https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai

2 https://www.consilium.europa.eu/en/policies/top-cyber-threats/

Intelligence and security organizations seem to be aware of this issue, which raises questions about whether and how they might deploy similar technologies in the delivery of security (Maclure, 2020; Hyder et al., 2026; Alfter et al., 2019). The Dutch government has published its *Wide Vision on AI,* outlining its foresight on harnessing AI and Generative AI (GenAI) for government services such as justice, defense, and security (Government of the Netherlands, 2024). Similarly, other NATO countries, such as the United States, suggest using Google Gemini to support defense and intelligence agencies in their data collection (Mitchell, 2024).

However, there are concerns about the deployment of AI models, such as GenAI systems, which require large datasets for training and task performance. The data may be collected from open-source sources or through web scraping, raising reasonable apprehensions about privacy, reliability, and (Fui-Hoon Nah et al., 2023; Mercer, 2023; Devanny and Dylan, 2023). We may be stuck in a dilemma regarding new data-driven technologies in intelligence and security investigations (Dórea, 2021; Yazici and Shayea, 2023).

How can we address these dilemmas in areas where the law is unclear or silent on emerging technologies for intelligence, security, and defense? (van der Constantino and van der Linden, 2025). Addressing dilemmas by public organizations, such as intelligence and security services, is vital in the context of accountable government and democracy (van Puyvelde, 2025; Bodó, 2022).

In earlier work, we articulated several principles that may help foster accountability practices in the face of intelligence and security challenges (Constantino, 2024). Yet it remains empirically unexplored how these principles are, or may be, operationalized, for instance, in the context of data-driven technologies in intelligence and security investigations (Constantino, 2024, p. 4, 5, 10–11, 18). Thus, our goals are to conduct empirical research to: (a) explore how intelligence and security actors perceive accountability in the context of AI deployment for digital investigations; and (b) explore to what degree the accountability principles proposed by Constantino and Wagner are potentially present among intelligence and security actors to navigate emerging technological challenges in data-driven investigations.

We have formulated the following research question to guide our research goals: How do democratic intelligence and security actors interpret and respond to accountability challenges posed by AI? Similarly, we formulated two sub-questions: SQ1: How do intelligence and security actors understand accountability in the wake of AI deployment? SQ2: To what extent are the principles of accountability from Constantino and Wagner present in human-AI interactions among intelligence and security actors? In section One, we provide an introduction. For context, in section Two, we present a background and literature review. Section Three outlines our methodology. In section Four, we present our findings. Section Five offers a discussion, and section Six provides conclusions.

This paper does not seek to develop a new theory of accountability. Rather, it offers an exploratory empirical contribution by examining how intelligence and security actors themselves understand accountability challenges arising from AI-supported investigations. The findings show that participants often located accountability concerns not in AI systems alone, but in the broader conditions under which such systems are developed, procured, governed, and used. Participants emphasized data acquisition, data sovereignty, AI reliability, explainability, human oversight, and dependency on external technological infrastructures. The article, therefore, contributes to debates on AI governance by showing how accountability in intelligence and security settings is operationalized through socio-technical practices under conditions of secrecy, uncertainty, and national security urgency.

We present empirical findings on the perspectives of various experts involved in intelligence and security work, whom we refer to as intelligence and security actors. This study largely collects perspectives from the Netherlands and triangulates its findings with those of other democratic NATO allies, including Belgium, Norway, and the UK. This work contributes to the academic community and to practitioners working in intelligence and security agencies, such as secret services, defense, Europol, NATO, and law enforcement, by providing knowledge on fostering accountability practices in the digital age. Likewise, this research is important to society at large because it examines issues related to fundamental rights in the digital age.

2 Background and literature review

This section develops the analytical lens for examining accountability in AI-supported intelligence and security investigations, identifying core topics relevant to our empirical analysis. We introduce "intelligence accountability" and its complexities, AI and Generative AI opportunities and challenges in intelligence and security work, and Constantino and Wagner's accountability principles.

2.1 Intelligence accountability and complexities

Accountability is commonly understood as a democratic mechanism through which actors may be required to explain and justify their conduct before a relevant forum (Schillemans, 2015; Bovens, 2010). In public administration, this is especially important when public organizations use technological tools to make or support decisions that affect citizens (Bodó, 2022). Likewise, intelligence and security services that deploy technological tools can be studied within the context of accountable government (Newberry, 2015; Walker and Archbold, 2020). Herein, we will refer to our topic as *intelligence accountability*.

Intelligence accountability is an opportunity to promote good governance, prevent abuse of power and contribute to the resilience of intelligence and security organizations in a democratic society (van Buuren, 2009; Wieringa, 2020; Koninkrijksrelaties, 2023).

However, as noted in Menkveld, intelligence accountability is complex due to the nature of intelligence and security work, exacerbated by the digital age that demands the use of data and technological tools, leading to unique challenges among democratic intelligence and security communities (Menkveld, 2021; Wessels and Schuilenburg, 2025).

As previously signaled, the first complexity concerns the nature of intelligence and security work. Vaage and Stenslie have outlined

that the primary duty or core task of intelligence and security practitioners is protecting national security from threat actors (Vaage, 2023). This primary task often involves various activities within the "intelligence cycle," including the collection of personal data, processing, analysis, and dissemination, both within and outside a country, to identify threats to a nation or its interests (Devanny and Dylan, 2023).[3]

The digital age has enabled intelligence and security agencies to collect intelligence through the collection of personal data from publicly available internet sources (OSINT), interception of telecommunications, and information sharing among intelligence and security partners (Scuro, 2026; Omand, 2021). Collecting data and information is necessary because it provides insights into threats or opportunities that might not be available elsewhere (Cayford, 2018). Processing and analysis are required to separate relevant information or data to convert it into an intelligence product (UK Ministry of Defence, 2023). According to Omand, analyzing data requires sensemaking or providing meaning to the data or information gathered; such information must be credible and actionable; otherwise, national security will be compromised (Omand, 2021; Omand in Dover et al., 2014). The graphic below illustrates, in a non-exhaustive manner, intelligence and security tasks, also known as the intelligence cycle (Figure 1).

Omand notes that gathering personal data may infringe on certain fundamental rights such as the right to privacy (Omand, 2021). Hence, intelligence and security organizations performing security tasks need to address questions regarding respect for human rights and the protection of the dignity and freedom of people exposed to intelligence and security investigations (Overeem in Kowalski, 2017).

Furthermore, Eldridge, Hobbs, and Moran argue that intelligence and security tasks cannot be performed easily or quickly by humans alone (Eldridge et al., 2017). Rather, it requires certain data-driven technologies to assist human practitioners in performing their tasks (Babuta et al., 2020), adding a new layer of complexity to intelligence accountability: practitioner-technology settings. Thus, the development and deployment of new data-driven technologies would affect how organizations and individuals perceive the right to privacy (Nissenbaum, 2009), urging organizations such as intelligence and security services to address potential challenges to privacy rights (Omand, 2021; Gürses and Hoboken, 2018).

Accordingly, we argue that the stakes are high, as the core duty of intelligence and security practitioners may be exacerbated by other factors that will challenge meaningful democratic intelligence accountability. The graphic below illustrates some of the complexities discussed above (Figure 2).

2.2 AI and generative AI systems in intelligence and security organizations

AI is a highly ambiguous and contested term, and no one can claim to have found a single definition (Ruschemeier, 2023). For instance, Schuett states that AI can refer to a range of systems, including those that generate text and make predictions (Schuett, 2021). McCarthy frames AI as a science and engineering of making "intelligent computer programs" seeking to simulate human intelligence, although AI is not biologically "intelligent" (McCarthy, 2007). Thus, as the purpose of this study is not to engage in technical arguments, but rather to explore intelligence accountability perspectives in human-machine interactions, we will follow the definition of Article 3(1) of the European Union (EU) Artificial Intelligence Act:

> *"AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"* (emphasis added).

Within the broad field of AI, we encounter Generative AI (GenAI), a subclass of machine learning that can assist with various other tasks such as generating images, text, or audio (Janjeva et al., 2023; Brynjolfsson et al., 2023; Walkowiak, 2023; Ritala et al., 2023).

2.2.1 Opportunities

Devanny argues that AI will play a vital role in intelligence and security cyberoperations, asserting that AI and GenAI models offer opportunities and advantages for intelligence and security organizations throughout the intelligence cycle (Devanny, 2024).

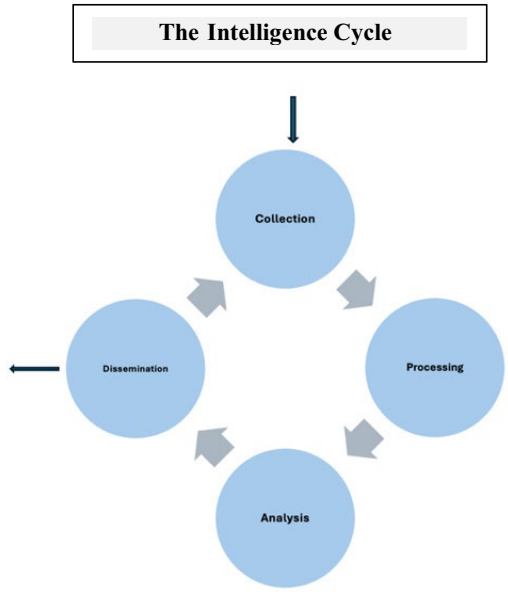

FIGURE 1
The intelligence cycle.

3 See also, https://www.dni.gov/index.php/what-we-do/what-is-intelligence.

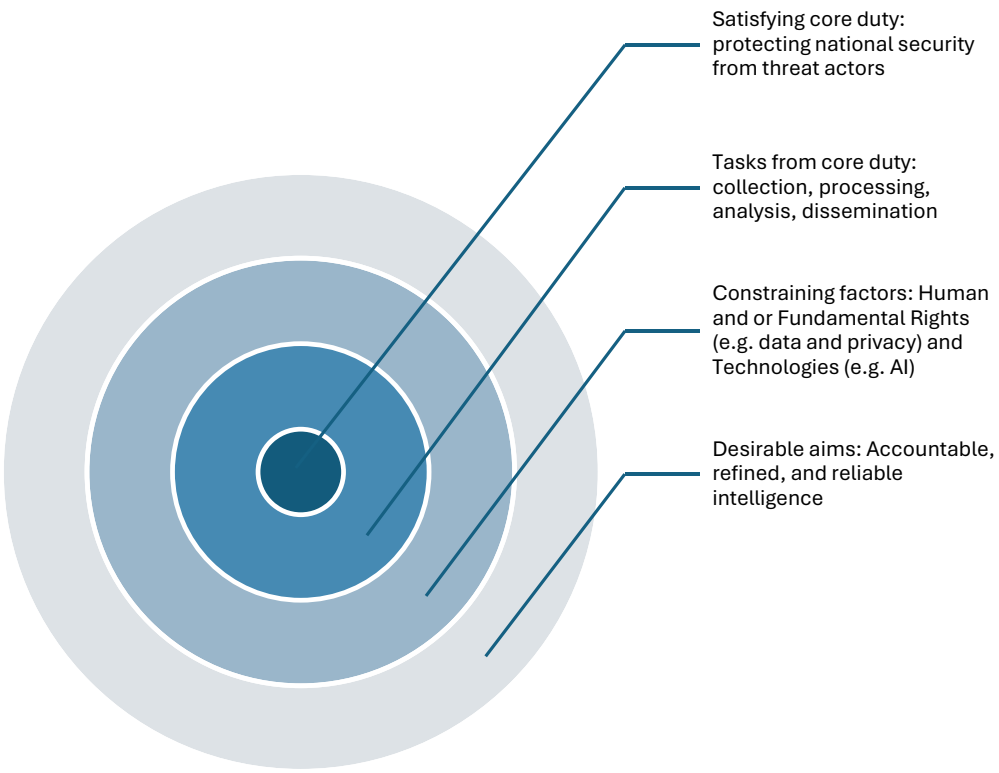

FIGURE 2
Layers of complexity in intelligence accountability.

2.2.1.1 Collection and processing

For instance, law enforcement and intelligence teams are adopting self-learning machine models, or so-called AI-enabled OSINT tools, for investigations involving structured and unstructured data (Europol, 2023). These AI-powered technologies, or LLM models, can connect to the internet, enabling the collection of information about current conflicts and other publicly available data at high speeds, allowing more efficiency in intelligence and security investigations (Ghioni et al., 2024). Suppose an analyst at the Dutch Intelligence and Security Services (the *AIVD*) wants to gain a deeper understanding of current political developments in Europe and the impact of the Russian-Ukrainian war on the Netherlands. An AI-enabled OSINT could help gather a diverse amount of information and data and provide an initial analysis, background information, and current trends on this conflict.

2.2.1.2 Analysis and dissemination

Hinton notes that GenAI could help quickly generate potential scenarios in which a conflict might play out (Hinton, 2023). Devanny et al. have established that GenAI in intelligence and security has the potential to significantly enhance threat intelligence analysis, for example, by identifying patterns or trends (Devanny and Dylan, 2023). GenAI may assist with tasks related to dissemination or communication (Devanny and Dylan, 2023). It can be used to generate intelligence reports and other forms of communication to inform decision-makers about their findings, or to generate information for strategic operations (Fui-Hoon Nah et al., 2023; Wahlstrom et al., 2022).

2.2.2 Challenges

AI systems pose diverse challenges from reliability and data training to overreliance and harm to human rights and fundamental rights, which prompts actors to exercise responsible behavior (Vogel et al., 2021; Devanny, 2024).

2.2.2.1 Reliability and accuracy

Reliability is of major importance for organizations under tremendous pressure, where mistakes can be unforgivable (Sutcliffe, 2011). Reliability speaks of the results that AI systems produce, whether it is replication over different times or inter-rate reliability (Kumar et al., 2026; Joachim et al., 2025; Farber, 2025). An AI system may be regarded as reliable based on the trustworthiness of its highly rated responses (Bianco et al., 2025). While accuracy measures the

correctness of an AI model's predictions, thereby supporting AI's trustworthiness (Constantino, 2025b).[4] Thus, accuracy may be understood as the validity, precision, or truthfulness of the model to produce predictions (Streiner, 2006).

AI models can face reliability and accuracy challenges due to insufficient or imprecise data training (Grote and Genin, 2024). Furthermore, AI systems' reliability is affected by hallucinations, false truths, false positives and false negatives of AI models, challenging the trustworthiness and readiness in high-stakes tasks such as cyber threat intelligence (Mezzi and Massacci, 2025; Mercer, 2023). Janjeva et al. highlight the difficulty for law enforcement practitioners to investigate child sexual abuse material due to the false positives and false negatives of AI-generated material, which challenges detection and response to crime (Janjeva et al., 2023). Thus, AI's unreliability can undermine national security and defense, leaving intelligence and security actors in a challenging position as they seek to fulfil their duties (Government of the Netherlands, 2024; Barzashka, 2023).

2.2.2.2 AI bias, historical data, black-boxes, and explainability

AI biases can emerge during the deployment stage of these models, for example, human bias leading to overreliance on the model,[5] and biases during the development stage, for instance, through historical data representation or stereotype reinforcement (Europol, 2025). As stated by Busuioc, bias can be hidden in historical datasets, misleading investigations and decision-making by unjustly targeting minorities or people with specific backgrounds (Busuioc, 2021). As noted by Taddeo et al., bias in human-AI interactions within intelligence and security communities poses a serious challenge to intelligence accountability (Taddeo et al., 2021).

Bathaee explains that AI systems become black boxes to humans because they rely on machine-learning techniques, making it hard for humans to audit and understand the model's behavior and outputs (Bathaee, 2018, p. 901). Thus, black boxes can challenge explainability in intelligence and security organizations (Blanchard and Taddeo, 2023, p. 12).

Explainable AI aims to make AI systems more understandable to humans, for instance, through documentation-based methods that allow humans to understand the AI's training data (Barredo Arrieta et al., 2019). Addressing explainable AI in security investigations is important because it allows the identification of causes of an AI system's low accuracy, also enabling practitioners and decision-makers to explain AI outputs, thereby increasing accountability in human-AI settings (Zocholl et al., 2025; Tanswell and Berg, 2025; Lapuschkin et al., 2019). It is worth noting that achieving explainability requires humans in the loop with sufficient AI literacy (Zocholl et al., 2025).

2.2.2.3 Overreliance and loss of skill

As signaled earlier, overreliance can lead to automation bias, making intelligence and security practitioners more likely to rely on automated outputs "despite warning signals or contradictory information from other sources" (Alon-Barkat, 2023). For instance, the well-documented Dutch childcare benefits scandal provides evidence of automation bias affecting the delivery of public services (Alon-Barkat, 2023).

The literature on professionalism and tacit knowledge offers insight into the challenges *of losing skills*. Tacit knowledge roughly refers to non-codified knowledge that is informally acquired via learning and experience and is rooted in action and routines, while explicit knowledge is usually expressed in formal language and easily recorded in different formats (Howells, 1996; Nonaka and von Krogh, 2009). In Sternberg, it is suggested that GenAI systems may lead to the loss of knowledge or skills if many human tasks are shifted to them (Sternberg, 2024). Likewise, we could argue that tacit knowledge becomes an important asset in intelligence and security work because non-codified knowledge cannot be transferred by machines to other machines or humans, thereby challenging an organization's ability to retain expert knowledge.

2.2.2.4 Privacy and other fundamental rights

Devanny notes that some GenAI systems use open-source data, raising concerns about data acquisition and questioning whether the data and information retrieved from these sources have been obtained lawfully (Devanny and Dylan, 2023, p. 5–7). Hinton addresses that respect for privacy and personal data protection is relevant because information from non-suspects may be compromised, leading to various privacy and data vulnerabilities (Hinton, 2023, p. 40).

Similarly, the Dutch government acknowledges that GenAI systems pose a challenge to privacy as a fundamental right due to the fact that GenAI systems are mostly trained with data from large-scale (web)scraping, public sources on the internet, or other digital sources (arguably data brokers), and data that may contain special personal information (Government of the Netherlands, 2024, p. 11). In the hypothetical scenario that the open-source intelligence was collected through illegal means (lack of proper consent or legitimate purpose justification), the outputs that serve to further train the GenAI model are likely to continue infringing fundamental rights, leading to public distrust regarding the legality and legitimacy of intelligence and security operations (Devanny, 2024, p. 7).

Further, GenAI tools, such as Gemini and OpenAI (ChatGPT), are exposed to cybercrime by various threat actors (Dave, 2024). The literature indicates that GenAI can be exploited by adversaries to threaten national security or the stability of the democratic legal order (Keast, 2023; Devanny, 2024). For example, Google's Gemini can track users' location, which can be detrimental to intelligence and security units if their tracked location is known by threat actors (Devanny and Dylan, 2023; Rogers, 2023).

Lastly, Zuiderveen-Borgesius highlights that algorithmic bias and back boxes pose a threat to an organization's accountability, as they affect fundamental rights, such as the right to non-discrimination (Zuiderveen Borgesius, 2020).[6] Sroka highlights that as black boxes limit explainability, the right to a fair trial and the possibility to contest AI outputs can be affected (Sroka, 2024). Similarly, Casaburo warns that AI systems used for security work may contain biases or historical

4 As the purpose of the paper is not to engage in technical terminologies, which are often contested in computer science, we will follow the scholarship that includes accuracy related to discussions concerning reliability (Vanbelle et al., 2025; Johnson, 2018).

5 See below at "*Overreliance*."

6 See also Europol (2025).

TABLE 1 AI opportunities, challenges, and consequences for intelligence and security.

Opportunities	Collection, processing, analysis, and dissemination
Challenges	(Un)reliability, Overreliance, loss of skill, Fundamental rights including breach of privacy, weak national security
Consequences	Good case scenario: Speed and capability enhancement in threat detection. Bad case scenario: AI dependency, unlawful fundamental rights breaches, weak security

data that lead to discriminatory results, potentially targeting vulnerable groups (Casaburo, 2024).[7]

We may summarize this section, as shown in Table 1, by stressing that, despite AI challenges, such as reliability, overreliance, and risks to security and fundamental rights, there are also opportunities, such as AI being used for the collection, analysis, or dissemination of information.

2.2.2.5 Governing AI in intelligence accountability

Vogel et al. found that intelligence practitioners have warned that AI systems may lead to analytical errors, potentially nudging them to make unwarranted assumptions about sources, data, and analysis, particularly with more junior practitioners, challenging the so-called "sensemaking" in intelligence and security operations (Vogel et al., 2021, p. 833). These challenges prompt urgent questions regarding meaningful human-AI governance in intelligence and security settings.

In Mäntymäki et al. and Birkstedt et al., it is suggested that AI governance is a system that implements rules, practices, and processes, along with technological tools, that align with an organization's strategies, values, and objectives to fulfil legal and ethical principles (Mäntymäki et al., 2022; Birkstedt et al., 2023). In this context, human actors are required to collaborate to exercise accountability to achieve meaningful human-AI governance oriented to protect citizens (Batool and Zowghi, 2025). These accountable practices may, for instance, entail ensuring responsible data acquisition practices, including collection and analysis, to identify potential risks or problems to citizens' rights (Zowghi and Rimini, 2023).

Currently, European intelligence and security agencies are not subject to the European AI Act, arguably leaving a (legal) accountability gap (van der Constantino and van der Linden, 2025). Therefore, the imperative for the literature to address accountability discussions that foster meaningful outputs for human-AI governance across the European intelligence and security sectors in the digital age.

AI governance implies attention to AI's robustness, accuracy, and reliability to ensure accountability (Batool and Zowghi, 2025; Constantino, 2025b). Addressing AI's reliability, whether in terms of accuracy or hallucinations, is crucial for understanding how AI outputs may fail humans and for implementing appropriate measures to counter these shortcomings (Vogel et al., 2021; Menkveld, 2021). Managing, reliable and accurate AI systems in intelligence and security investigations is needed in order to protect society's human or fundamental rights from harmful AI systems (Constantino, 2025b; European Union Agency for Fundamental Rights, 2019; Selçuk and Kurt Konca, 2025).

Johnson highlights that, despite the opportunities AI systems offer to support decision-making in high-stakes environments such as military operations, their deployment challenges human agency through automation bias and over-reliance (Johnson, 2026).[8] Likewise, limited AI awareness, along with vague or non-existent guidelines regarding meaningful human oversight in intelligence and security communities, can undermine accountability in human-AI settings (Wessels and Schuilenburg, 2025; Zocholl et al., 2025; Soares and Grimmelikhuijsen, 2024). In turn, if these socio-technical challenges are not properly addressed, they risk eroding accountability within intelligence and security communities; thus, the importance of humans-in-the-loop being empowered to exercise effective control in AI-supported decision-making (Johnson, 2026). In this context, Bode et al. suggest that governance arrangements need to address humans-in-the-loop as the ones retaining legal and moral accountability (Bode et al., 2025).

The literature on algorithmic accountability increasingly suggests that accountability in AI-supported environments cannot be understood solely through traditional legal responsibility frameworks. While traditional accountability scholarship emphasizes answerability, responsibility, and democratic control (Schillemans, 2015; Bovens, 2010), recent work on AI governance shows that accountability also depends on the interaction between human actors, organizational rules, data infrastructures, technical systems, and institutional safeguards (Batool and Zowghi, 2025; Birkstedt et al., 2023; Mäntymäki et al., 2022). In this sense, accountability in AI-supported settings is not only a question of who is legally responsible after a decision has been made, but also how organizations govern data acquisition, AI trustworthiness, explainability, human oversight, and the institutional conditions under which AI outputs are produced and used (Bode et al., 2025; Zocholl et al., 2025; Zowghi and Rimini, 2023). This reveals a tension in the literature. On the one hand, some scholarship notes that accountable AI primarily may be framed or understood through technical requirements such as robustness, accuracy, reliability, explainability, and transparency (Zocholl et al., 2025; Constantino, 2025b; Barredo Arrieta et al., 2019). On the other hand, broader AI governance scholarship emphasizes that accountability cannot be reduced to technical design because AI systems operate within socio-technical environments shaped by human judgment, organizational practices, legal mandates, institutional cultures, data governance arrangements, and power relations (Constantino, 2025a; Batool and Zowghi, 2025; Birkstedt et al., 2023; Bodó, 2022; Mäntymäki et al., 2022; Busuioc, 2021). This distinction is particularly important for intelligence and security organizations, where AI systems are not deployed in neutral settings but within complex environments shaped by secrecy, urgency, uncertainty, national security mandates, and fundamental rights constraints (Wessels and Schuilenburg, 2025; Vogel et al., 2021; Menkveld, 2021; Omand, 2021, 2024).

The literature on AI governance and responsible data practices suggests that accountability gaps may also arise earlier in the AI lifecycle, including through unlawful or inappropriate data acquisition,

7 Compare above with *"AI bias, historical data, black-boxes and explainability."*

8 See also above at section 2.3.

historical bias in training data, insufficient human oversight, and dependency on external technological infrastructures (Europol, 2025; Devanny and Dylan, 2023; Zowghi and Rimini, 2023). This means that accountability, in the context of AI, cannot be treated solely as an algorithmic problem, but rather as a broader governance issue encompassing data, infrastructures, human factors, and other institutional arrangements.

Yet the literature remains unexplored on how intelligence and security practitioners themselves interpret accountability when AI systems enter their complex environments. This paper addresses that gap by exploring how democratic intelligence and security actors understand accountability in AI-supported investigations and by assessing how these perspectives extend the accountability principles proposed by Constantino and Wagner.

2.3 Accountability principles in intelligence and security domains

In Constantino and Wagner (Constantino, 2024, p. 7–8), we identified several principles relevant to intelligence and security work, including *acting within duty,* which refers to operating strictly within the limits of the legal mandate or legal duty. Exp*lainability* suggests that intelligence and security practitioners have an obligation to justify their decisions, particularly when working with data and algorithmic tools. *Necessity* requires practitioners to justify interference with fundamental rights, while *proportionality* requires actions to be proportional to the threat. R*ecord-keeping* maintains detailed documentation of operations, while *Reporting provide*s stakeholders with information on intelligence and security practices. *Redress* is framed as a means to amend errors. *Continuous Independent Oversight* is a highly contested topic that is often described as constant or near real-time end-to-end supervision.[9] These principles serve as a sensitizing lens for analyzing our empirical findings, which examine intelligence accountability challenges in the context of AI in intelligence and security investigations.

3 Methodology

3.1 Research design

Bovens argues that accountability can best be studied through a qualitative approach because it is a social phenomenon that cannot be quantified but rather requires an understanding of its interpretation and social context (Bovens, 2010). Consequently, this study adopts a qualitative approach to conduct explorative empirical research focusing on intelligence and security accountability perspectives in the context of AI deployment for digital investigations. As well as understanding how and to what extent the principles of accountability outlined by Constantino and Wagner are reflected in the practices of intelligence and security organizations when deploying AI systems for intelligence investigations. Thus, prescribed by Bryman, semi-structured interviews were conducted to enable participants to express their own viewpoints rather than present the author's perspective (Bryman, 2016). Likewise, we conducted a literature review to gain a deeper understanding of the current state of AI and Generative AI in relation to intelligence, security, and defense operations (von Soest, 2023; Denzin, 2017).[10]

3.2 Data collection

Von Soest highlights the importance of adding expert interviews since they are the observers and the mechanisms of policy, allowing the opportunity to understand how "x" interacts with "y" (von Soest, 2023). This approach allows establishing a connection between prior knowledge and data collected (Bryman, 2016; Jones et al., 2010).

We conducted purposive sampling of experts to allow information richness in this complex topic (Schreier, 2018). As such, we reached out to more than fifty (50) potential participants from NATO allies, including Belgium, the United Kingdom, Norway, and the Netherlands, receiving the most positive responses from participants with Dutch representation. In total, we have interviewed twenty-two (22) participants over three (03) months. According to Guest et al., saturation can be achieved as early as six interviews, when no new information can be obtained through further data collection (Guest and Bunce, 2006). However, there is no consensus on the exact number of interviews required to reach saturation.

Our selection criteria for participants were based on countries with similar democratic intelligence accountability and oversight frameworks (Born and Leigh, 2005; Gill, 2020). To avoid bias, our "Dutch" responses were triangulated with those from participants representing Norway, the United Kingdom, and Belgium (von Soest, 2023; Bowen, 2009; Coleman, 1958). Similarly, to avoid potential "echo chambers," we combined practitioners from different organizations and perspectives, all with knowledge and or experience in topics related to data-driven technologies, AI, and intelligence and security investigations (Modgil et al., 2024; Braun, 2023; Robinson, 2014; Noy, 2008). Our approach provides richness and diversity of perspectives suitable for explorative qualitative analysis (Charmaz, 2021; Corbin and Strauss, 2014).

Some participants decided to respond individually, while others provided joint statements, as shown in our findings section. Our research had ethics approval from the TU Delft Board of Ethics. The questions presented during the interviews were open-ended, designed to stimulate the conversation and allow participants to share their views (Jones et al., 2010). The questions were developed based on discussions with our supervising team and on prompts from the existing literature (Bryman, 2016).

Three (03) pilot interviews with participants from Australia, the Netherlands, and the United Kingdom helped us refine our interview protocol and hypothetical case scenario for actors with a more technical background.[11] Following Rosson and Carroll in Sears and Jacko (2003), we designed a hypothetical case study on deploying GenAI systems in intelligence and security contexts to stimulate discussion of our topic. The case scenario was constructed based on current literature.

9 *Continuous Independent Oversight* is the subject of a separate study in our forthcoming work.

10 See below at section 4 "AI-GenAI systems for intelligence and security."

11 See below "hypothetical case scenario."

TABLE 2 Overview of participants.

Participant ID	Jurisdiction	Professional background	Interview format
P1	Netherlands	Intelligence and security oversight	Audio recording
P2	Netherlands	Academia intelligence and security studies	Audio recording
P3	Netherlands	Intelligence and/or security practice	Audio recording
P4	Netherlands	Data protection oversight	Audio recording
P5	Netherlands	National defense practice	Audio recording
P6	Netherlands	Strategic analyst AI-military	Audio recording
P7	Netherlands	Academia intelligence and security studies	Audio recording
P8	Netherlands	Digital rights advocate	Audio recording
P9	Netherlands	Academia privacy and digital rights	Audio recording
P10	Netherlands	Academia privacy and digital rights	Audio recording
P11	Netherlands	Academia intelligence and security studies	Audio recording
P12	Netherlands	Digital rights advocate	Audio recording
P13	Netherlands	Political actor	Audio recording
P14	Belgium	Academia intelligence and security studies	Audio recording
P15	Netherlands	Intelligence and/or security practice	Handwritten
P16	Netherlands	Intelligence and/or security practice	Handwritten
P17	Netherlands	Intelligence and/or security practice	Handwritten
P18	Netherlands	Intelligence and/or security practice	Handwritten
P19	Netherlands	Intelligence and/or Security Practice	Handwritten
P20	Netherlands	Intelligence and security oversight	Handwritten
P21	United Kingdom	Academia intelligence and security studies	Audio recording
P22	Norway	National defense practice	Audio recording

Hypothetical Case Scenario: AI, Threats, and security.

Sky is a security and intelligence analyst. Today, Sky received intelligence indicating that a malicious actor is preparing a sophisticated cyberattack within the country. The malicious actor is leveraging a cutting-edge criminal Generative AI (GenAI) model, known as *WormX*, trained on leaked darknet data to conduct deep reconnaissance, actively mining leaked employee credentials, technical documentation, and public infrastructure datasets to build detailed profiles of personnel and identify vulnerabilities in the energy and healthcare sectors. It remains unclear whether the threat actor is operating independently or as part of a coordinated effort involving state or non-state entities.

At the same time, Sky is offered access to a GenAI-powered OSINT model, called *GenMax*. It is developed by a third party and capable of accelerated data collection and analysis, arguably enabling Sky to gain awareness and identify the malicious actor. However, *GenMax* has raised concerns regarding **privacy and data protection**, **flawed recommendations,** and unclear **accountability mechanisms**. Time of the essence to prevent critical disruption of energy and healthcare systems. **How might Sky proceed in resolving this scenario?**

3.3 Data analysis

Interview recordings from participants were transcribed for coding and quote purposes. We utilized various recording methods, including Cisco Webex, Voice Memos, Teams, and Zoom. Unfortunately, most of participant P13's interview is missing due to technical issues. The interviews lasted an average of ninety (90) to one hundred twenty (120) minutes. Due to national security and clearance measures, some interviews were handwritten; these interviews were used as verbatim quotes. Thus, to avoid misrepresentation of participants' statements, we shared a draft of the quotes used in this work with them. Similarly, given the specialized and sensitive nature of the field, identifying personal and location-specific details has been withheld to protect participant confidentiality. This approach is coherent with social science literature and legal frameworks regarding participants' privacy (Crow and Wiles, 2008; Thorne, 1980). For the reader's benefit, we present below a table providing a general overview of the participants included in this study (Table 2).

The semi-structured interview transcripts were analyzed in ATLAS.ti using an inductive thematic analysis approach. Thematic analysis is flexible and can be used in different contexts, which involves extracting key themes that emerge from the data collected by the researcher (Martínez-Buelvas et al., 2025; Braun, 2023; Braun, 2006; Driessen et al., 2024; Phillips and Hardy, 2002; Bryman, 2016).

Charmaz suggests that themes are essentially the same as codes and that some grounded theory criteria can be applied in thematic analysis (Charmaz, 2021). Given the exploratory and interpretive nature of this qualitative research, we acknowledge the possibility of researcher bias in data interpretation and theme construction. Thus, to mitigate this risk, the coding process involved iterative comparison across interviews, memo writing, triangulation across participants

from different professional sectors and jurisdictions, and ongoing discussion within the research team regarding theme development and interpretation. Additionally, quotations used in this study were shared with participants where appropriate to reduce risks of misrepresentation. The resulting codes were clustered into broader categories and subsequently developed into higher-order themes, including AI-Data Governance Challenges, AI Trustworthiness Challenges, Humans-in-the-loop Factors, and In-house Resources. The quotations were selected as illustrative examples of our findings and themes, which explain participants' perspectives (Eldh and Årestedt, 2020; O'Brien et al., 2014).

Similarly, in practice, thematic saturation was assessed iteratively throughout the coding process (Guest and Bunce, 2006). As interviews progressed, recurring patterns emerged regarding AI governance gaps, reliability concerns, explainability limitations, and human oversight across participants from diverse professional backgrounds and jurisdictions. After approximately the twelfth interview, additional interviews produced limited, substantially new thematic insights, suggesting sufficient thematic depth for the exploratory purposes of this study.

We acknowledge that a limitation of this article may be the anonymization of our sources and the use of handwritten recording methods. Furthermore, while this study includes perspectives from multiple NATO allies, the findings remain primarily grounded in Dutch institutional and legal contexts. Accordingly, the findings should not be interpreted as universally representative of all intelligence and security organizations. Rather, the study provides exploratory insights into how accountability challenges surrounding AI are currently perceived among democratic intelligence and security actors operating within relatively comparable accountability frameworks. In future research, the pool of participants could be expanded to include other NATO countries, with particular interest in the current geopolitical context and the future of the Rule of Law and the International Legal Order.

We note other limitations of this study: participants devoted most of their time to discussing AI-Data challenges, with very little attention to AI opportunities. Similarly, participants have lightly addressed principles of accountability such as *reporting*, *record keeping*, *proportionality,* and *necessity* (Constantino, 2024). These principles are still to be empirically studied in more depth.

4 Findings

We present detailed research findings on the analysis of two main themes in this study: "AI Challenges on Intelligence Accountability" and "Intelligence Accountability Perspectives."[12]

4.1 AI challenges on intelligence accountability

AI Challenges on Intelligence Accountability include (a) AI-Data governance Challenges; (b) AI Trustworthiness Challenges; (c) Humans-in-the-loop Factors; (d) In-house Resources.

4.1.1 AI-data governance challenges

When asked about the potential risks or challenges of AI to intelligence accountability,[13] we found that most of our participants were less concerned about AI itself. For instance, P1 said that the main issue, instead of AI, "*is more about big datasets.*"[14] Participants' responses were more focused on data governance challenges and current AI gaps, such as the lack of AI frameworks governing intelligence and security organizations. Participants identified data acquisition in the context of collection and use of historical data, the impact on human and fundamental rights, and data and technology sovereignty, as challenges inherent to what we named "*AI-Data Governance Challenges*."

4.1.1.1 Lack of AI frameworks

Participants' statements illustrate the pressing need for legal frameworks governing AI in intelligence and security services. For example:

> "*I think it does not really matter which type of AI system a security and intelligence analyst uses here, because from a legal perspective, there's no policy in place there, so whatever is going to happen, then let us say there is a violation because there's personal data of a person being used, which should not be used. Then it's without a legal basis because there is no clear legal framework; in a sense, the State is acting arbitrarily, and then you do not even have to go into a discussion of necessity and proportionality because it's an arbitrary behaviour which then infringes on fundamental rights.*" (P9).

> "*If you want to make an analysis of a bulk data set and use AI to ask the AI, please make an analysis and who are the terrorists in this data set? Then you have like a big thing, and you should create new legal norms for that.*" (P7).

However, Dutch intelligence and security practitioners provided alternative ways in which potential use of AI may be examined, for example, having regard to data management practices or even ethical decisions:

> "*The Act, Wiv 2017, is precise about data requirements, no matter how you get the data… So under [the] current framework, AI is not prohibited.*" (P17, P18, and P19).

> *"It can be helpful as well to work on your ethical competencies to make the right decision and to think about what is right or wrong, perhaps also in a situation where new circumstances arise, new technologies arise, and the law might be outdated on that point or might not give enough guidance on what to do. Then it could be*

12 The other main themes concerning our data analysis will be addressed in two separate papers concerning intelligence oversight and a proposal to manage the challenges and complexities from intelligence accountability in the digital age.

13 Note that as the question was explorative in nature, and some participants were assisted with our hypothetical case scenario, some participants decided to focus more on GenAI, while others dismissed the GenAI suggestion to instead address on broad terms AI, which can include GenAI aspects, in accordance with the definition of the European AI Act, Article 3(1): "*AI system means a machine-based system that is designed to operate with varying levels of autonomy and… infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*"

14 Please refer to Table 1: Overview Participants' Demographics.

> *helpful to reflect on it from an ethical perspective… which does not diminish the role of law, but can add additional support."* (P3).

4.1.1.2 Data acquisition

Although data acquisition or collection is a clear legal task concerning the intelligence cycle, we found that P10 and P1 showed concerns about "overenthusiastic" practices in collecting data, challenging the basis for data acquisition:

> "*Since there's more data that they [intelligence and security practitioners] can find, I think there's a risk that they may be over enthusiastic in collecting too much because, yeah, it may be helpful someday, and we saw that, for instance, in the Netherlands, where the army started collecting loads of data online and monitoring Twitter accounts, etcetera. Well, they did not have, to begin with, the legal basis to do so, and I trust that they were well-meaning, but yeah, I find that problematic.*" (P10).

> "*With digital investing data-driven investigations, you have several bulk datasets and you look for information in these several bulk datasets and yeah, you have to think about. How do you acquire the data? Is that lawful? But also how do you process that data*?" (P1).

4.1.1.3 Data and technology sovereignty

Participants highlighted that another challenge posed by AI systems in intelligence and security work is the absence of mechanisms to ensure that data and technology remain sovereign:

> *"If that technology is off the shelf, so maybe it's not being developed as a sovereign technology. What mechanisms are in place to ensure that actually the other country that owns the data, but might not own the company, that data can remain sovereign. Do we have the right laws and governance mechanisms in place to ensure that data is sovereign?"* (P22).

> "*The problem with using external LLMs, generative AI, or other technologies is that if we do not control the data and the model, then we will have issues with reliability and analysis.*" (P15 and P16).

> "*I think what they [intelligence and security communities] are mostly interested in, also in managing their own resources, is to make sure that whatever they do and interact with the system, stays with them, so that they are not training a model that is not infrastructure, which can be compromised by parties that they do not want to have access to information that you know there's a whole lot of things, to make sure that what happens there stays there just from an infrastructure and organisational perspective.*" (P9).

4.1.2 AI trustworthiness challenges

Participants addressed algorithmic bias, historical data, black boxes, reliability, including accuracy and hallucinations, and the explainability of AI models as a challenge to intelligence accountability. We framed this section as "*AI Trustworthiness Challenges*."

4.1.2.1 Algorithmic bias

Participants have suggested that biases integrated into AI systems are a problem for intelligence and security accountability because they can lead to false positives and false negatives, negatively affecting the protection of fundamental rights and security:

> "*machines are biased because the algorithms have been shaped by humans. So it is like, even the weakness of the machines are due to the humans.*"[15] (P14).

> "*You get all these kinds of profiles, and with the use of these profiles, people who should not have been included or have not done anything wrong will be included in search programmes or whatever, and then people who actually might commit a crime but are not included in a profile because they are white or living in another neighbourhood they are not seen at all, so it's not working from both sides.*" (P12).

> "*In a positive scenario, person A is bad, and if the computer says 'no threat', then that's a huge problem, because the computer misleads. How do you make it accountable? A machine cannot be held accountable, but people we are accountable, people cannot be replaced by a technology that takes over our accountability.*" [16] (P16).

4.1.2.2 Historical data

When reflecting on cases of AI used by law enforcement and intelligence agencies, P12 highlights that historical data can lead to discrimination:

> "*when it comes to the use of algorithms or AI by the police, it is historical data… you will find a lot of cases in which people with a migrant background or people with colour or people within certain neighbourhoods in the Netherlands are being stopped or are being arrested or being suspect of a criminal offences*" (P12).

While participant P22 suggested that intelligence and security operations cannot be solely based on historical data:

> "*If you think that the best way to train a model is on all your historical data of military maneuvers or military activity. Well, that does not necessarily tell you how to conduct future operations. Because it's all historical data, is not it? And you do not necessarily want to use that as your basis for future operations.*" (P22).

15 See below under discussion, the interconnection between human and technical problems.

16 Statements from P16 also reveal further problems such as AI's accuracy or reliability which will be discussed below.

4.1.2.3 Black boxes

Participants addressed black box challenges in intelligence and security accountability, highlighting the limitations in explaining decisions:

> "*if you ask the model a question, and the answer comes from the technology itself, which is a black box. So this kind of problem affects explainability and reliability in human-AI interaction.*" (P16).

> "*the same problem, compounded by the kind of black box, and then it has consequences for real people. But these real people have a big problem fighting whatever injustice can be done against them because they do not know what happened because of this black box.*" (P8).

4.1.2.4 Reliability

Participants' statements regarding AI's hallucinations and accuracy are linked to reliability challenges, highlighting the consequences to intelligence accountability:

> "*you have to ensure that the intelligence information is credible, and if you use AI you need to focus a lot on accountability and just to make sure that you have the right information which is not like made by AI hallucinations or other AI algorithms that you cannot trust.*" (P7).

> "*These [LLMs] outputs are to be checked by the practitioner, not for drawing conclusions, because the technology is new and can hallucinate and fill gaps. So LLMs or hypothetically GenAI cannot be used for too high-level questions, as they are not good at answering fundamental questions. So, you have to be careful about what and where you use it for.*"[17] (P15).

> "*…things like ChatGPT, they put things down very authoritatively… So this is an issue of accuracy and this is an issue of just like truthfulness.*" (P6).

Lastly, P22 provided a rather pessimistic prediction about whether these technical challenges may ever be solved:

> "*big companies that are developing, for example, these language models. Every time they think that they are making headway, that the language model will not, for example, hallucinate or will not, for example, do this or behave in this way, they have not solved the problem yet.*" (P22).

4.1.3 Humans-in-the-loop factors

Another issue we encountered was the role of intelligence practitioners as willing humans-in-the-loop to counter or minimize the above challenges from AI-Data Governance or AI Trustworthiness. Humans-in-the-loop factors present AI awareness or literacy, and the risks of relying on technology to offload work.

4.1.3.1 AI awareness and critical thinking

Participants have stressed the need for practitioners' technological awareness:

> "*AI is a tool to help with data processing. Thus, we need to know the type AI and be aware of false positives and false negatives… know how to use AI; knowing the AI techniques used to develop the system… knowing what the system does.*" (P17, P18, P19).

> "*If the tools do not give you the probability, but simply say this is the outcome, I mean, you have to take into account yourself that such tools are fabulating.*" (P2).

We observed that participants hinted at human agency in the context of accountable practitioners (humans-in-the-loop):

> "*Part of the aspect of intelligence accountability is ethics in the use of technology. For example, humans give computers instructions, humans have the role of ethical sensemakers… In our work, humans have to come up with conclusions and assessments, that is also part of intelligence accountability.*" (P16).

> "*humans in the loop are part of the duty of care when working with AI. There should always be a human control.*" (P17, P18, P19).

Similarly, P8 and P15 stated that humans-in-the-loop need to apply critical thinking against "seductive" AI systems:

> "*Sometimes there is too much trust in technology, and it's like, yeah, well, the system says that, the system gives this person a high risk score… without looking at why this system is pointing out that person, for example, or critically reflecting.*" (P8).

> "*careful how you use the LLM model because the results of the model look perfect, even the quality looks true, and the layout looks true. So be careful because the model can be seductive to humans.*" (P15).

4.1.3.2 Overreliance and AI offloading

We found that participants such as P6, P8, and P16 shared their concerns about the risks of AI overreliance and using AI to offload work:

> "*A good practitioner needs to be as correct as possible; that is why humans should not rely on technology… Relying on the model, there is a possibility of a shortcut to the truth.*" (P16).

> "*most people are just using it [AI] to try and ease off the load of, like, their standard work, like the work that they already have, and not necessarily use it as a tool for learning, but more just for offloading.*" (P6).

17 Note that this finding also relates to see below at 'humans-in-the-loop factors, particularly in intelligence and security work where it is argued that AI cannot replaced the complex and refined work of intelligence and security agencies.

> "*it can be fairly relaxing if somebody else or something else tells you what to do because you do not have to think about it yourself. But then you can also question the value of this human in the loop.*" (P8).

We observed that among intelligence practitioners, academics, and civil society, the concern is mutual about AI awareness and the vital role of intelligence practitioners as accountable humans-in-the-loop:

> "*so hopefully for the next decade or two, analysis remains with humans. It can be machine-supported, but it's about engaging with and working for a consumer in the dissemination…at the end of the day, it's like intelligence services and agencies are there for humans by humans.*" (P14).

4.1.4 In-house resources

Participants P2, P6, P14, and P22 all addressed the lack of resources and dependency on external providers, which can open back doors and challenge intelligence accountability:

> "*if you look at the Netherlands, a small community, a few thousand people with a small budget, you know, there's no comparison, the Dutch intelligence budget, is in the hundreds of millions, I imagine, not much more. Compared that to the budget of, I do not know, Twitter or Elon Musk.*" (P14).

> "*the secret community is now still doing a few things themselves, but it's likely dependent of what happens in the public and in the commercial sector… So, engaging with third-party providers would depend on the level of trust.*" (P2).

> "*geopolitics and the supply chain of AI, it's just so international… to develop one system, so many different countries are involved with phases… You have chips from Taiwan… and then they have to fly to the United States. And then they end up in Germany, back in Europe, like that kind of thing. And at any point, there could be a back door that is built in.*" (P6).

> "*we are connecting our operational technology with our information technology, which means that potentially someone can hack a dam… That's pretty important, for national security.*" (P22).

4.2 Intelligence accountability perspectives

Participants provided their views regarding intelligence work and satisfying accountability,[18] thereby allowing us to identify two major perspectives: (a) security and fundamental rights dilemmas; and (b) intelligence accountability in a digital age.[19]

4.2.1 Security and fundamental rights dilemmas

Participants shared their feelings and perspectives regarding the dilemma to protect national security while caring for human and fundamental rights:

> "*Our work is full of complexities, we have the duty of care to assess probabilities of attacks, human factors, assess national security, preserve the democratic order, and protect the security of the state, so everything we do is an assessment.*" (P17, P18, P19).

> "*Who's going to be held responsible if *** goes wrong?. When we speak about national security, mistakes are not fine… In my philosophy, we are trying to do the right thing… We are trying to prevent disaster from happening… So you have a duty.*" (P5).

> "*…the interests of an intelligence or intelligence service are. Let us say they are in a tense relationship with fundamental rights… it's a great paradox.*" (P11).

> "*I think it's important to expect priority of rights.*" (P13).

> "*Sometimes, that is, we cannot have pure security, or we cannot have maximum security and maximum human rights respect at the same time. It's always in wherever we strike the balance.*" (P10).

4.2.2 Intelligence accountability in a digital age

There has been a broad discussion among participants about what is expected and what it means to be an accountable intelligence practitioner in a digital age defined by data and AI.[20]

We observed that participants made a link between practitioners' values and legal duties regarding the choices they make about data and technology in intelligence and security investigations:

> "*…there are always choices about what kind of data is gathered… So if we have different values, then this is only exacerbated by the use of technology.*" (P8).

> "*In the service, we have a duty of care, for example… we need to be careful with all data acquired during the processing… requests us to process data in the correct manner.*" (P17).

> "*for example, GenAI for OSINT would require applying the duty of care and creating a set of rules or safeguards for humans.*" (P19).

> "*Conducting impact assessments is part of our duty of care, so before buying an AI system we need to go through the explainability of the system, checking for outcomes of algorithms and AI, impact on consequences on human rights, assessment if AI will achieve the goals of the organisation; assess how would you use it to check how important or will affect intelligence investigations.*" (P17, P18, P19).

18 Compare above with section 2.2.

19 Due to time constraints and the amounts of data from our interviews, in this paper we only focus on those complexities related to the technology and security and fundamental rights perspectives. In a separate paper we cover other complexities such as regulation constraining intelligence and security work; or power dynamics including external oversight and political interference challenging the fulfillment of meaningful intelligence accountability.

20 Compare above with "Humans-in-the-loop Factors."

> "*if you deal with national security or defence, you have to take measures to keep the information protected and secure because you are dealing with state secrets.*" (P1).

> "*…where is the data stored? How is it stored? How is it protected?. where are your servers located and how are they protected*?" (P6).

Further information emerged regarding intelligence accountability in the digital age. For instance, when asked, participants about what can be done in cases of misuse of data or AI tools which may be detrimental to national security and fundamental rights, we received various responses such as *"immediate sanctions"* or *"interventions"* to delete data unlawfully collected:

> "*there should also be immediate sanctions and measures. If something is not right.*" (P5).

> "*intervention[s] and the power to delete data.*" (P1).

> "*unlawfully detained data, it should be deleted.*" (P8).

> "*Reporting and record-keeping internally is good to provide reasons of proportionality. Record keeping and recording is good because we can look back at why sometimes, or trace back decisions.*" (P17, P18, P19).

5 Discussion

We interpret the findings of the previous sections in light of the background provided in section 2, including Constantino and Wagner's conceptual framework on intelligence accountability. In particular, accountability appears increasingly challenged through socio-technical interactions involving data governance arrangements, AI reliability assessments, limitations in explainability, and human oversight capacities. AI systems appear to intensify the practical importance of principles such as acting within duty and explainability, while simultaneously exposing practical accountability gaps concerning data sovereignty, algorithmic opacity, and overreliance on automated outputs. Our findings, therefore, extend Constantino and Wagner's framework by illustrating that accountability in AI-supported intelligence environments is not exercised solely through legal compliance, but also through continuous socio-technical judgment regarding data quality, AI trustworthiness, and meaningful human oversight.

Accordingly, the findings suggest that accountability in AI-supported intelligence investigations is not merely a question of assigning legal or moral responsibility. Rather, accountability increasingly depends upon the continuous capacity of practitioners and organizations to critically evaluate, govern, supervise, and contextualize AI-generated outputs within broader approaches, including democratic, moral, and legal safeguards.

The findings also engage with broader debates concerning the so-called "responsibility gap" in AI governance. Although participants acknowledged that AI systems may generate unreliable, biased, or opaque outputs, they consistently rejected the idea that accountability could be delegated to machines themselves. Instead, participants reaffirmed the continuing responsibility of human practitioners and organizations to exercise judgment, verify AI's outputs and intervene.

In this respect, the findings strongly suggest that intelligence and security practitioners remain accountable regardless of the technology in use, though, due to the nature of technological development in the digital age, this accountability is challenged by various socio-legal-technical factors.

5.1 About AI-data governance gaps

Consistent with the literature, we found that participants, such as P9, are aware that the lack of legal frameworks governing AI systems in intelligence and security organizations is a challenge to intelligence accountability in the execution of their tasks (van der Constantino and van der Linden, 2025). However, participants have suggested that, as AI deployment is not currently prohibited, AI use may therefore be governed by current data and algorithmic management arrangements, imposing a legal duty on practitioners to handle data with care. We observe that at least for Dutch practitioners, the focus of their accountability is centered on data practices, currently legislated under the *Wiv 2017*, regardless of the technology. This perspective also reveals their awareness of the importance of upholding accountability principles, such as "acting within duty."

Additionally, to address current legal gaps in AI deployment within intelligence communities, it may be helpful to reflect on and implement ethical practices that support intelligence accountability, as P3 suggests. This perspective moves beyond legal accountability. Rather, this view is more consistent with a moral or ethical side of accountability, which seeks to hold public servants accountable (van Puyvelde, 2025; Constantino, 2024, p. 19–21; Bovens, 2010; van Buuren, 2009).

We identify that AI governance also includes data acquisition practices, with participants' perspectives stressing "how" and "what" data is acquired. The digital age allows for harnessing large data streams, which can be helpful to intelligence and security work (Cayford, 2018; Omand, 2024). However, as our participants suggested, it is important not to be "overenthusiastic" about data acquisition or processing, and to consider "*necessity*" and "*proportionality*" principles.

Concerns related to data and technology sovereignty affecting intelligence accountability are also part of AI-Data Governance Challenges. Participants have stressed that, in intelligence and security work, it is necessary to preserve data generated by AI models built by external actors to ensure that the data and the technology remain sovereign. Data and technology sovereignty may be undermined by technological and financial constraints, pressing intelligence and security organizations to outsource technology to other actors, challenging the integrity of national security secrets and accountability (P6, P22). These concerns align with the literature, which suggests that intelligence work is so delicate that, in most cases, it requires secrecy to protect sensitive information, or else national security may be compromised (Bellaby, 2021). Thus, preserving data and technology sovereignty in the undertaking to protect national security and other fundamental rights at stake suggests being part of intelligence accountability.

In anticipation of our next section, it is worth noting that, despite boxing our themes as "*AI challenges*," "*Trustworthiness Challenges*," or

"*Human Factors*," these challenges are not mutually exclusive. Rather, as our interviews reveal, these challenges are intertwined. Thus, one of the most significant findings of this study is that participants frequently framed accountability challenges less as problems originating from AI itself and more as problems rooted in data governance practices surrounding AI systems, such as bulk data acquisition, historical datasets, legality of collection practices, and effective human oversight. This finding is important because it suggests that accountability debates in intelligence and security contexts should not narrowly focus on algorithmic systems in isolation. Rather, AI accountability appears deeply interconnected with broader questions concerning data governance, lawful data acquisition, storage practices, infrastructure dependency, and effective human control over information flows. Accordingly, our findings suggest that intelligence accountability in the context of AI systems supporting investigations may be better understood as a broader socio-legal-technical governance problem rather than solely a technical problem of algorithms.

5.2 About AI trustworthiness challenges

We observe that AI Trustworthiness requires addressing problems arising from algorithmic bias, historical data, hallucinations, black boxes, accuracy and reliability, which will challenge intelligence accountability. Consistent with the literature, the responses from participants P14 and P8 indicate that algorithmic bias is not only a machine-inherent problem but also a human one, underscoring that human bias can be replicated across the digital ecosystem (Europol, 2025). As observed by our participants, machine bias can be a problem when the algorithms produce incorrect results, for instance, leading to false positives or false negatives in investigations. Using historical data to train a model to predict future operations can deceive and weaken intelligence and security organizations and may harm the fundamental rights of innocent people, undermining intelligence accountability (Busuioc, 2021; Casaburo, 2024; Taddeo et al., 2021).

AI's black boxes also challenge accountability for intelligence by undermining practitioners' ability to provide meaningful explanations of AI outputs and organizational decisions (Sroka, 2024; Zocholl et al., 2025). Thus, from our participants' responses, we identify that the principle of "exp*lainability" appears necessary and relevant to intelligence accountability when working with AI tools* (Constantino, 2024).

The above-discussed challenges amount to AI systems being regarded as unreliable or untrustworthy (Bianco et al., 2025), which, of course, is not very beneficial to intelligence and security work, particularly when the stakes are high and mistakes are unforgivable (Sutcliffe, 2011). AI trustworthiness issues cannot be confined to the technology itself, as if the "machine" were to be held accountable (Constantino, 2025a). Rather, the discussion naturally leads to human factors considerations.

5.3 About humans-in-the-loop factors

Participants' remarks about practitioners needing AI awareness, including knowledge of AI techniques and an understanding of AI systems' capabilities, are consistent with the literature on AI explainability, suggesting the need for human overseers who can effectively understand AI's outputs (Zocholl et al., 2025).

Moreover, meaningful humans-in-the-loop are those humans who retain agency and do not delegate it to an AI system (P6). These perspectives are consistent with the literature's warnings against overreliance and the call on humans-in-the-loop to exercise meaningful agency (Bode et al., 2025; Johnson, 2026). Participants have noted that relinquishing human agency risks missing out on truths, challenging the practitioners' responsibilities to be "as correct as possible" (P16) so as to prevent disastrous effects to national security (Vogel et al., 2021; Menkveld, 2021). In turn, intelligence accountability for *"acting within duty" may be challenged if human practitioners delegate or relinquish their agency to AI*.

Further, effective humans-in-the-loop in intelligence and security settings are vital because intelligence practitioners are the only ones with the capacity to provide sensemaking; machines cannot do so (P15). This approach is consistent with the literature regarding intelligence work (UK Ministry of Defence, 2023; Omand, 2021). Thus, we observe the necessity of retaining and encouraging effective humans-in-the-loop in intelligence and security communities, as they have the skills to deliver refined intelligence outputs (P14). We must remain aware that, when shifting tasks from humans to AI systems, we may risk losing refined expert knowledge (Howells, 1996; Nonaka and von Krogh, 2009; Sternberg, 2024).

Lastly, we observe that humans-in-the-loop who do not possess AI literacy, or who constantly rely on AI outputs, risk being unable to provide explanations in the course of their undertaking, in such cases this limitation can challenge people's fundamental rights who have the right to an explanation about how or why they ended up being investigated by intelligence and security services (Alon-Barkat et al., 2025; Sroka, 2024).

5.4 About intelligence accountability perspectives

Participants have addressed broader perspectives on accountability work, highlighting the dilemmas involved in everyday intelligence and security practice, such as making assessments in the best interests of national security and fundamental rights, while recognizing that mistakes are not acceptable in their work (P17, P18, P19, P5). These dilemmas are consistent with the literature acknowledging complexity in high-stakes environments (Omand, 2021; Overeem in Kowalski, 2017; Babuta et al., 2020; Menkveld, 2021; Vaage, 2023). These broader perspectives suggest that practitioners' duties are not confined to national security but also extend to protecting fundamental rights. These perspectives suggest that intelligence practitioners are always faced with a tension between national security and fundamental rights (P11), so there cannot be a "*maximum security and maximum human rights at the same time*" (P10). Rather, it will require making assessments to find an appropriate balance that can satisfy these necessities.

In a narrower sense, intelligence accountability in the digital age makes us aware that practitioners and organizations will have to make choices based on their values, and technology use will play a role in these decisions, ultimately with consequences on national security and society (P8; Mäntymäki et al., 2022; Birkstedt et al., 2023; Burdon, 2019). For example, OSINT tools leveraging AI will require the creation of frameworks to safeguard humans, which is part of the duty of care of intelligence and security organizations (P19).

Thus, we can observe that the intelligence accountability perspectives described by participants as a complex dilemma between choices and consequences can be linked to the literature on accountability, which suggests a constant interaction between "actor" and "forum," resulting in constant explanations for decision-making (Bovens, 2010). Similarly, when addressing complexities in intelligence work, we identify at least two layers of complexity in (Dutch) intelligence accountability: (a) the inherent duty to protect national security from threat actors; (b) the duty to manage data and technological deployment in a proper manner, for instance, in a way that does not unlawfully violate fundamental rights. That is not to say that these layers of complexity are exhaustive.[21]

Based on our participants' accountability perspectives, we note that "*redress*," "*reporting,*" and "*record keeping*" appeared in the interviews but were less dominant than the other principles discussed above (Constantino, 2024). For instance, "*redress*" may be identified as "*sanctions and measures if something is not right" (P5), or* "*data destruction*" in cases of "*unlawfully detained data" (P8).* On the other hand, "*reporting*" and "*record keeping*" are seen as enablers that require practitioners to motivate and trace decisions (P17, P18, P19).

5.5 Final remarks

It is worth referring to "*In-house Resources*" for two main reasons. Firstly, if AI tools are outsourced to third parties, effective or meaningful intelligence accountability may be constrained because intelligence organizations may be subjected to third parties' willingness to foster accountability. This matter is worth further empirical study. Secondly, if the aim is to foster intelligence accountability, we are inclined to think that regulating third parties is also appropriate; whether the European AI Act will do so remains to be tested. For now, it may be preferable to identify feasible solutions that would allow intelligence communities to open black boxes, enabling, for instance, the application of explainability principles to foster intelligence accountability.

In turn, this reflects that intelligence accountability, in the context of our topic, is complex because it depends on a range of factors, such as socio-legal and socio-technical arrangements across the entire AI chain. Importantly, accountability does not shift from humans to machines. Rather, the findings reinforce that human actors remain legally and morally accountable, while evolving technologies such as AI systems will challenge the operational conditions under which accountability must be exercised. From our research findings and analysis, we observe two main underlying constraints in meaningful intelligence accountability, which are related to the literature about socio-technical systems (Patriarca et al., 2021), that is:

a. Human Factors: meaning AI literacy and willingness to perform human-in-the-loop tasks.
b. Technology Constraints: including data-AI governance, technology trustworthiness, and financial resources to keep in-house technology

These socio-technical challenges to intelligence accountability are not to be treated or read in isolation. These challenges also require attention to legal frameworks. Thus, we may argue about socio-legal-technical challenges in intelligence and security accountability.

6 Conclusion

Our goal was to conduct exploratory research to understand intelligence and security actors' perspectives on accountability in the digital age, with particular emphasis on AI. Some participants opted to address AI in general terms, while others focused on more specific examples, such as GenAI or OSINT AI-powered. We submit that the examples do not alter our findings on intelligence accountability perspectives in relation to human-machine interaction, as the study is not centered on computer science technicalities.

We found that almost all our participants expressed concern about data-AI governance, challenging national security and the protection of fundamental rights. Despite the lack of a current tailored AI framework, some intelligence actors suggested that no matter the technology, the core issue is data governance shaping AI development and deployment. Participants strongly identified data management practices and technology use as closely linked to their duty of care. Although, as noted above, AI systems offer opportunities to intelligence and security work, participants addressed AI trustworthiness challenges, including reliability, accuracy, bias, black-box models, historical data, and hallucinations. Moreover, participants have addressed Humans-in-the-loop Factors as both potential constraints and enablers of intelligence accountability, highlighting practitioners' duty to serve as critical sense-makers. In any case, participants seem to agree that intelligence practitioners have a duty of care regarding data handling and the use of technology during their investigations. This view supports our argument that agency should remain with humans and that accountability cannot be shifted from humans to machines or algorithms.

Overall, participants addressed intelligence accountability in the digital age as a complex task that requires assessing components such as protecting national security, proper data and technology management, and consideration of fundamental rights. In relation to Constantino and Wagner's principles of accountability, we found that participants' perspectives were heavily focused on "acting within duty" and "explainability" (Constantino, 2024). Other accountability principles, such as "proportionality" or "necessity," were scarcely present during the interviews.

Intelligence accountability is a socio-legal-technical problem which requires policy considerations. For instance, implementing a holistic review of the current arrangements, including accountability of intelligence practitioners, oversight bodies, private actors, and state actors, to protect societal interests. Future work can investigate options and formulate strategies to manage these socio-legal-technical challenges.

In answering our research questions, we found that democratic intelligence and security actors interpret and respond to accountability challenges posed by AI tools in much the same way. Despite our participants' diverse backgrounds, their responses indicated that intelligence accountability, in the context of AI, relates to how practitioners use and manage data and technological tools to navigate data, while paying attention to core duties to protect national security and uphold human and fundamental rights. Thus, our semi-structured interviews revealed that AI itself does not represent a challenge to "*intelligence accountability*." Rather, intelligence accountability in the digital age extends to issues such as data governance, technological

21 See above at 2.1 "Intelligence and Security Complexities."

trustworthiness, human factors, and adequate resources. Ultimately practitioners are accountable, yet socio-legal-technical constraints will challenge effective intelligence accountability in the digital age.

Data availability statement

The datasets presented in this article are not readily available due to the nature of the research and the requirement of consent from the participants. Requests to access the datasets should be directed to the corresponding author/s.

Ethics statement

Ethical approval was obtained from the TU Delft Ethics Board. Written informed consent to participate in this study was obtained from the participants.

Author contributions

JC: Supervision, Writing – review & editing, Conceptualization, Methodology, Software, Investigation, Writing – original draft, Formal analysis, Visualization, Project administration, Data curation, Validation, Resources. FK: Writing – review & editing. SG: Writing – review & editing, Validation, Formal analysis, Methodology. BW: Writing – review & editing, Funding acquisition, Formal analysis, Methodology, Supervision, Validation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by NWO project (no. KICH1.VE01.20.004).

Acknowledgments

We thank Mark Burdon from Queensland University of Technology for his time and constructive feedback.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Alfter, B., Muller-Eiselt, R., and Spielkamp, R. (2019). Automating Society, AlgorithmWatch. Available online at: https://algorithmwatch.org/en/automating-society/ (Accessed June 11, 2024).

Alon-Barkat, S. (2023). Human–AI interactions in public sector decision making: "automation bias" and "selective adherence" to algorithmic advice. *J. Public Adm. Res. Theory* 33, 153–169. doi: 10.1093/jopart/muac007

Alon-Barkat, S., Busuioc, M., Schwoerer, K., and Weißmüller, K. S. (2025). Algorithmic discrimination in public service provision: understanding citizens' attribution of responsibility for human versus algorithmic discriminatory outcomes. *J. Public Adm. Res. Theory* 35, 469–488. doi: 10.1093/jopart/muaf024

Babuta, A., Oswald, M., and Janjeva, M. (2020). Artificial Intelligence and UK National Security: Policy Considerations. Royal United Services Institute, p. 57. Available online at: https://www.rusi.orghttps://www.rusi.org (Accessed January 28, 2026).

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2019). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *arXiv*. doi: 10.48550/arXiv.1910.10045

Barzashka, I. (2023). Seeking strategic advantage: the potential of combining artificial intelligence and human-centred wargaming. *RUSI J.* 168, 26–32. doi: 10.1080/03071847.2023.2282862

Bathaee, Y. (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. Cambridge, Massachusetts: Harvard Journal of Law and Technology.

Batool, A., and Zowghi, D. (2025). AI governance: a systematic literature review. *AI Ethics* 5, 3265–3279. doi: 10.1007/s43681-024-00653-w

Bellaby, R. (2021). "Intelligence and the just war tradition, the need for a flexible ethical framework," in *National Security Intelligence and Ethics*, (London: Routledge).

Bianco, G. L., Lo Bianco, G., Cascella, M., Li, S., Day, M., Kapural, L., et al. (2025). Reliability, accuracy, and comprehensibility of AI-based responses to common patient questions regarding spinal cord stimulation. *J. Clin. Med.* 14:1453. doi: 10.3390/jcm14051453

Birkstedt, T., Minkkinen, M., Tandon, A., and Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Res.* 33, 133–167. doi: 10.1108/INTR-01-2022-0042

Blanchard, A., and Taddeo, M. (2023). The ethics of artificial intelligence for intelligence analysis: a review of the key challenges with recommendations. *Digit. Soc.* 2:12. doi: 10.1007/s44206-023-00036-4

Bode, I., Nadibaidze, A., Watts, T., and Zhang, Q. (2025). Ensuring the exercise of human agency in AI-based military systems: concerns across the lifecycle. *Ethics Inf. Technol.* 27:50. doi: 10.1007/s10676-025-09861-2

Bodó, B. (2022). Maintaining trust in a technologized public sector. *Polic. Soc.* 41, 414–429. doi: 10.1093/polsoc/puac019

Born, H., and Leigh, H. (2005). 'Making Intelligence Accountable: Legal Standards and Best Practice for Oversight of Intelligence Agencies - GSDRC'. Available online at: https://gsdrc.org/document-library/making-intelligence-accountable-legal-standards-and-best-practice-for-oversight-of-intelligence-agencies/ (Accessed February 12, 2026).

Bovens, M. (2010). Two concepts of accountability: accountability as a virtue and as a mechanism. *West Eur. Polit.* 33, 946–967. doi: 10.1080/01402382.2010.486119

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qual. Res. J.* 9, 27–40. doi: 10.3316/QRJ0902027

Braun, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Braun, V. (2023). Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. *Int. J. Transgender Health* 24, 1–6. doi: 10.1080/26895269.2022.2129597

Bryman, A. (2016). Social Research Methods. 4th ed. Oxford University Press. Available online at: https://www.deslegte.com/social-research-methods-73544/ (Accessed March 25, 2025).

Brynjolfsson, E., Li, D., and Raymond, D. (2023). Generative Ai at Work. Rochester, NY: Social Science Research Network. Available online at: https://papers.ssrn.com/abstract=4426942 (Accessed February 11, 2026).

Burdon, M. (2019). The significance of securing as a critical component of information security: an Australian narrative. *Comput. Secur.* 87:101601. doi: 10.1016/j.cose.2019.101601

Busuioc, M. (2021). Accountable artificial intelligence: holding algorithms to account. *Public Adm. Rev.* 81, 825–836. doi: 10.1111/puar.13293

Casaburo, D. (2024). Ensuring fundamental rights compliance and trustworthiness of law enforcement AI systems: the ALIGNER fundamental rights impact assessment. *AI and Ethics* 4, 1569–1582. doi: 10.1007/s43681-024-00560-0

Cayford, M. (2018). The effectiveness of surveillance technology: what intelligence officials are saying. *Inf. Soc.* 34, 88–103. doi: 10.1080/01972243.2017.1414721

Charmaz, K. (2021). The pursuit of quality in grounded theory. *Qual. Res. Psychol.* 18, 305–327. doi: 10.1080/14780887.2020.1780357

Coleman, J. S. (1958). Relational analysis: the study of social organizations with survey methods. *Hum. Organ.* 17, 28–36. doi: 10.17730/humo.17.4.q5604m676260q8n7

Constantino, J. (2024). Accountability and oversight in the Dutch intelligence and security domains in the digital age. *Front. Polit. Sci.* 6:1383026. doi: 10.3389/fpos.2024.1383026

Constantino, J. (2025a). "Accountable AI: it takes two to tango," in *The Digital Decade*, (Berlin: Nomos).

Constantino, J. (2025b). Accuracy, Robustness and Cybersecurity. In: The EU Artificial Intelligence (AI) Act: A Commentary Wolters Kluwer. Eds. C. N. Pehlivan, N. Forgó and P. Valcke. The Netherlands: Wolters Kluwer.

Corbin, J., and Strauss, J. (2014). Basics of Qualitative Research. SAGE Publications, Inc. Available online at: https://uk.sagepub.com/en-gb/eur/basics-of-qualitative-research/book235578 (Accessed February 12, 2026).

Crow, G., and Wiles, G. (2008). 'Managing Anonymity and Confidentiality in Social Research: The Case of Visual Data in Community Research'. Available online at: https://www.researchgate.net/publication/279480811_Managing_anonymity_and_confidentiality_in_social_research_the_case_of_visual_data_in_Community_research (Accessed February 12, 2026).

Dave, P. (2024). 'US National Security Experts Warn AI Giants Aren't Doing Enough to Protect Their Secrets', Wired Available online at: https://www.wired.com/story/national-security-experts-warn-ai-giants-secrets/ (Accessed June 11, 2024).

Denzin, N. K. (2017). *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: Routledge.

Devanny, J. (2024). Artificial Intelligence and Cyber Power: Foreign Policy Implications. Miami, FL: Jack D. Gordon Institute, Florida International University. 39.

Devanny, J., and Dylan, H. (2023). Generative AI and intelligence assessment. *RUSI J.* 168, 16–25. doi: 10.1080/03071847.2023.2286775

Dórea, F. C. (2021). Data-driven surveillance: effective collection, integration, and interpretation of data to support decision making. *Front. Vet. Sci.* 8:633977. doi: 10.3389/fvets.2021.633977

eds. R. Dover, M. Goodman and C. Hillebrand (2014). *Routledge Companion to Intelligence Studies*. Milton Park, Abingdon, Oxon: Routledge.

Driessen, H. P. A., Bakker, E. M., Rietjens, J. A. C., Luu, K. L. N., Lugtenberg, M., Witkamp, F. E., et al. (2024). A qualitative study on redefining normality in relatives of patients with advanced cancer. *Cancer Med.* 13:e7211. doi: 10.1002/cam4.7211

Eldh, A. C., and Årestedt, L. (2020). Quotations in qualitative studies: reflections on constituents, custom, and purpose. *Int J Qual Methods* 19:1609406920969268. doi: 10.1177/1609406920969268

Eldridge, C., Hobbs, C., and Moran, M. (2017). Fusing algorithms and analysts: open-source intelligence in the age of "big data". *Intelli National Sec* 33, 391–406. doi: 10.1080/02684527.2017.1406677

European Union Agency for Fundamental Rights (2019). Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights. Available online at: https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect (Accessed February 12, 2026).

Europol (2023). 'The Second Quantum Revolution – The Impact of Quantum Computing and Quantum Technologies on Law Enforcement'. Europol Innovation Lab observatory report, Publications Office of the European Union, Luxembourg. Available online at: https://www.europol.europa.eu/publication-events/main-reports/ai-and-policing (Accessed October 28, 2025).

Europol (2025). *AI Bias in Law Enforcement - A Practical Guide*. Luxembourg: Europol Innovation Lab observatory report, Publications Office of the European Union.

Farber, S. (2025). Comparing human and AI expertise in the academic peer review process: towards a hybrid approach. *High. Educ. Res. Dev.* 44, 871–885. doi: 10.1080/07294360.2024.2445575

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., and Chen, L. (2023). Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *J. Inf. Technol. Case Appl. Res.* 25, 277–304. doi: 10.1080/15228053.2023.2233814

Ghioni, R., Taddeo, M., and Floridi, L. (2024). Open source intelligence and AI: a systematic review of the GELSI literature. *AI & Soc.* 39, 1827–1842. doi: 10.1007/s00146-023-01628-x

Gill, P. (2020). Of intelligence oversight and the challenge of surveillance corporatism. *Intell. Natl. Secur.* 35, 970–989. doi: 10.1080/02684527.2020.1783875

Government of the Netherlands. (2024). 'The Government-Wide Vision on Generative Al of the Netherlands'. The Ministry of the Interior and Kingdom Relations. Available online at: https://www.government.nl/documents/parliamentary-documents/2024/01/17/government-wide-vision-on-generative-ai-of-the-netherlands (Accessed June 3, 2024).

Grote, T., and Genin, K. (2024). Reliability in machine learning. *Philos Compass* 19:e12974. doi: 10.1111/phc3.12974

Guest, G., and Bunce, A. (2006). How many interviews are enough?: an experiment with data saturation and variability. *Field Methods* 18, 59–82. doi: 10.1177/1525822X05279903

Gürses, S., and van Hoboken, J. (2018). Privacy after the Agile Turn. In: The Cambridge Handbook of Consumer Privacy. Cambridge Law Handbooks. Cambridge University Press. Eds. E. Selinger, J. Polonetsky and O. Tene. 579–601.

Hinton, P. (2023). Generative AI and wargaming: what is it good for? *RUSI J.* 168, 34–41. doi: 10.1080/03071847.2023.2282863

Howells, J. (1996). Tacit knowledge. *Technol. Anal. Strateg. Manag.* 8, 91–106. doi: 10.1080/09537329608524237

Hyder, Z., Keng, S., and Fiona, N. (2022). "Artificial Intelligence, Machine Learning, and Autonomous Technologies in Mining Industry." In *Research Anthology on Cross-Disciplinary Designs and Applications of Automation, edited by Information Resources Management Association*, Hershey, PA: IGI Global Scientific Publishing. 478–492. doi: 10.4018/978-1-6684-3694-3.ch024

Janjeva, A., Alexander, H., Sarah, M., Alexander, K., and Anna, G. (2023). The Rapid Rise of Generative AI: Assessing Risks to Safety and Security United Kingdom Centre for Emerging Technology and Security 87. Available online at: https://cetas.turing.ac.uk/publications/rapid-rise-generative-ai (Accessed June 03, 2024).

Joachim, M. V., Dodson, T. B., and Laviv, A. (2025). How artificial intelligence differs from humans in peer review. *J. Oral Maxillofac. Surg.* 83, 1040–1050. doi: 10.1016/j.joms.2025.03.015

Johnson, M. S. (2018). Measures of agreement to assess attribute-level classification accuracy and consistency for cognitive diagnostic assessments. *J. Educ. Meas.* 55, 635–664. doi: 10.1111/jedm.12196

Johnson, J. (2026). Can AI behave ethically during military crises? Preserving human moral agency. *Int. Aff.* 102, 63–83. doi: 10.1093/ia/iiaf191

Jones, I. R., Leontowitsch, M., and Higgs, M. (2010). The experience of retirement in second modernity: generational habitus among retired senior managers. *Sociology* 44, 103–120. doi: 10.1177/0038038509351610

Keast, J. (2023). Shadow Play Australia Australian Strategic Policy Institute. Available online at: http://www.aspi.org.au/report/shadow-play (Accessed June 12, 2025).

Koninkrijksrelaties, M. B. Z. (2023). AIVD/MIVD 2018–2023 - Verslag van het Functioneren van de Diensten - Rapport - Rijksoverheid.nl. The Hague: Ministerie van Algemene Zaken. Available online at: https://www.rijksoverheid.nl/documenten/rapporten/2023/09/01/verslag-van-het-functioneren-van-de-diensten-aivd-mivd-2018-2023 (Accessed September 6, 2023).

Kowalski, M. (2017). Ethics of Counterterrorism. The Hague. Available online at: https://www.boomgeschiedenis.nl/product/100-568_Ethics-of-counterterrorism (Accessed June 1, 2023).

Kumar, A., Poungpeth, N., Yang, D., Farrell, E., Lambert, B. L., and Groh, M. (2026). When large language models are reliable for judging empathic communication. *Nature Machine Intelligence* 8, 1–13. doi: 10.1038/s42256-025-01169-6

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., and Müller, K.-R. (2019). Unmasking clever Hans predictors and assessing what machines really learn. *Nat. Commun.* 10:1096. doi: 10.1038/s41467-019-08987-4

Maclure, J. (2020). The new AI spring: a deflationary view. *AI Soc.* 35, 747–750. doi: 10.1007/s00146-019-00912-z

Mäntymäki, M., Minkkinen, M., Birkstedt, T., and Viljanen, M. (2022). Defining organizational AI governance. *AI Ethics* 2, 603–609. doi: 10.1007/s43681-022-00143-x

Martínez-Buelvas, L., Rakotonirainy, A., Grant-Smith, D., and Oviedo-Trespalacios, O. (2025). Policy considerations for the safe and equitable integration of connected and automated vehicles. *Case Stud. Transp. Policy* 22:101631. doi: 10.1016/j.cstp.2025.101631

McCarthy, J. (2007). What is Artificial Intelligence?. Available online at: http://www-formal.stanford.edu/jmc/ (Accessed January 28, 2026).

Menkveld, C. (2021). Understanding the complexity of intelligence problems. *Intell. Natl. Secur.* 36, 621–641. doi: 10.1080/02684527.2021.1881865

Mercer, S. (2023). 'Welcome to Willowbrook: The Simulated Society Built by Generative Agents'. United Kingdom. Available online at: https://cetas.turing.ac.uk/publications/welcome-willowbrook-simulated-society-built-generative-agents (Accessed June 3, 2024).

Mezzi, E., and Massacci, F. (2025). "Large language models are unreliable for cyber threat intelligence," in *Availability, Reliability and Security*, ed. M. Dalla Preda (Cham: Springer Nature Switzerland), 343–364.

Mitchell, B. (2024). 'Google Gets Authorization to Work with Top-Secret Intelligence, Defense Data. Available online at: https://fedscoop.com/google-gets-authorization-to-work-with-top-secret-intelligence-defense-data/ (Accessed June 11, 2024).

Modgil, S., Singh, R. K., Gupta, S., and Dennehy, D. (2024). A confirmation Bias view on social media induced polarisation during Covid-19. *Inf. Syst. Front.* 26, 417–441. Available at:. doi: 10.1007/s10796-021-10222-9

Newberry, S. (2015). Public sector accounting: shifting concepts of accountability. *Public Money Manag.* 35, 371–376. doi: 10.1080/09540962.2015.1061180

Nissenbaum, H. (2009). Privacy in Context Stanford University Press. Available online at: https://www.sup.org/books/law/privacy-context (Accessed February 11, 2026).

Nonaka, I., and von Krogh, G. (2009). Perspective—tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. *Organ. Sci.* 20, 635–652. doi: 10.1287/orsc.1080.0412

Noy, C. (2008). Sampling knowledge: the hermeneutics of snowball sampling in qualitative research. *Int. J. Soc. Res. Methodol.* 11, 327–344. doi: 10.1080/13645570701401305

O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., and Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Acad. Med.* 89, 1245–1251. doi: 10.1097/ACM.0000000000000388

Omand, D. (2021). How Spies Think. Available online at: https://www.penguin.co.uk/books/312746/how-spies-think-by-omand-david/9780241385197 (Accessed March 13, 2025).

Omand, D. (2024). Examining the ethics of spying: a practitioner's view. *Crim. Law Philos.* 18, 805–818. doi: 10.1007/s11572-023-09704-5

Patriarca, R., Falegnami, A., Costantino, F., Di Gravio, G., De Nicola, A., and Villani, M. L. (2021). WAx: an integrated conceptual framework for the analysis of cyber-socio-technical systems. *Saf. Sci.* 136:105142. doi: 10.1016/j.ssci.2020.105142

Phillips, N., and Hardy, N. (2002). *Discourse Analysis*. Thousand Oaks: SAGE Publications, Inc.

Ritala, P., Ruokonen, M., and Ramaul, L. (2023). Transforming boundaries: how does ChatGPT change knowledge work? *J. Bus. Strateg.* 45, 214–220. doi: 10.1108/JBS-05-2023-0094

Robinson, O. C. (2014). Sampling in interview-based qualitative research: a theoretical and practical guide. *Qual. Res. Psychol.* 11, 25–41. doi: 10.1080/14780887.2013.801543

Rogers, R. (2023). 'How to Stop Google Bard from Storing your Data and Location', Wired. Available online at: https://www.wired.com/story/google-bard-location-data-tracking-ai/ (Accessed June 10, 2024).

Ruschemeier, H. (2023). AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. *ERA Forum* 23, 361–376. doi: 10.1007/s12027-022-00725-6

Schillemans, T. (2015). Managing public accountability: how public managers manage public accountability. *Int. J. Public Adm.* 38, 433–441. doi: 10.1080/01900692.2014.949738

Schreier, M. (2018). "Sampling and generalization," in *The SAGE Handbook of Qualitative Data Collection. Edited Uwe Flick*, (London, United Kingdom: SAGE Publications Ltd).

Schuett, J. (2021). Defining the Scope of AI Regulations. Rochester, NY: Social Science Research Network.

Scuro, V. (2026). "Open-source intelligence (OSINT) for researchers and practitioners," in *Researching a Rigged Game: Digital Approaches to Tracing the Illicit Trade in Cultural Objects*, eds. E. Smith and S. Austin (Cham: Springer Nature Switzerland), 11–28.

eds. A. Sears and J. A. Jacko (2003). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Third Editiion*. Mahwah, N.J: CRC Press.

Selçuk, S., and Kurt Konca, N. (2025). AI-driven civil litigation: navigating the right to a fair trial. *Comput. Law Secur. Rev.* 57:106136. doi: 10.1016/j.clsr.2025.106136

Soares, C., and Grimmelikhuijsen, S. (2024). Screen-level bureaucrats in the age of algorithms: an ethnographic study of algorithmically supported public service workers in the Netherlands police. *Inf. Polity* 29, 277–292. doi: 10.3233/IP-220070

Sroka, T. (2024). 'Artificial intelligence and the right to a fair trial', Artificial Intelligence and International Human Rights Law. Edward Elgar Publishing, pp. 250–277. Available online at: https://www.elgaronline.com/edcollchap-oa/book/9781035337934/book-part-9781035337934-22.xml (Accessed March 25, 2026).

Sternberg, R. J. (2024). Do not worry that generative AI may compromise human creativity or intelligence in the future: it already has. *J. Intelligence* 12:69. doi: 10.3390/jintelligence12070069

Streiner, D. L. (2006). "Precision" and "accuracy": two terms that are neither. *J. Clin. Epidemiol.* 59, 327–330. doi: 10.1016/j.jclinepi.2005.09.005

Sutcliffe, K. M. (2011). High reliability organizations (HROs). *Best Pract. Res. Clin. Anaesthesiol.* 25, 133–144. doi: 10.1016/j.bpa.2011.03.001

Taddeo, M., McNeish, D., Blanchard, A., and Edgar, E. (2021). Ethical principles for artificial intelligence in national defence. *Philos. Technol.* 34, 1707–1729. doi: 10.1007/s13347-021-00482-3

Tanswell, F. S., and Berg, F. S. (2025). The Philosophical Prospects of Large Language Models in the Future of Mathematics. Available online at: https://philsci-archive.pitt.edu/27017/ (Accessed March 23, 2026).

Thorne, B. (1980). "You still takin' notes?" fieldwork and problems of informed consent*. *Soc. Probl.* 27, 284–297. doi: 10.2307/800247

UK Ministry of Defence (2023). 'Joint Doctrine Publication 2-00, Intelligence, Counter-Intelligence and Security Support to Joint Operations', (4). Available online at: https://assets.publishing.service.gov.uk/media/653a4b0780884d0013f71bb0/JDP_2_00_Ed_4_web.pdf (Accessed January 28, 2026).

Vaage, B. H. (2023). How good is Norwegian intelligence? *Int. J. Intell. Counterintell.* 36, 968–979. doi: 10.1080/08850607.2021.1986792

van Buuren, J. (2009). Secret Truth. The EU Joint Situation Centre Amsterdam Eurowatch. Available online at: https://www.statewatch.org/media/documents/news/2009/aug/SitCen2009.pdf (Accessed December 05, 2024).

Constantino, J. van der, and van der Linden, J. (2025). 'Artificial intelligence (AI) systems not covered by the AI regulation', *The European Artificial Intelligence Act*. Cham: Springer Nature. 211–235.

van Puyvelde, D. (2025). The rise of open-source intelligence. *Eur. J. Int. Secur.* 10, 530–544. doi: 10.1017/eis.2024.61

Vanbelle, S., Engelhart, C. H., and Blix, E. (2025). Measuring agreement in diagnostics: a practical guide for researchers. *Stat. Med.* 44:e70299. doi: 10.1002/sim.70299

Vogel, K. M., Reid, G., Kampe, C., and Jones, P. (2021). The impact of AI on intelligence analysis: tackling issues of collaboration, algorithmic transparency, accountability, and management. *Intell. Natl. Secur.* 36, 827–848. doi: 10.1080/02684527.2021.1946952

von Soest, C. (2023). Why do we speak to experts? Reviving the strength of the expert interview method. *Perspect. Polit.* 21, 277–287. doi: 10.1017/s1537592722001116

Wahlstrom, A., Revelli, A., Riddell, S., Mainor, D., and Serabian, R. (2022). 'The IO Offensive: Information Operations Surrounding the Russian Invasion of Ukraine'. Available online at: https://cloud.google.com/blog/topics/threat-intelligence/information-operations-surrounding-ukraine (Accessed June 12, 2024).

Walker, S., and Archbold, S. (2020). *The New World of Police Accountability*. Thousand Oaks, CA: SAGE Publications.

Walkowiak, E. (2023). *Generative AI and the Workforce: What Are the Risks?* Rochester, NY: Social Science Research Network.

Wessels, M., and Schuilenburg, M. (2025). Algorithmic accountability and the use of real-time AI tools: an ethnography of police officers' daily use of AI violence detection in public spaces. *Eur. J. Policing Stud.* 8, 220–244. doi: 10.5553/EJPS.000043

Wieringa, M. (2020). 'What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY: Association for Computing Machinery. 1–18.

Yazici, İ., and Shayea, I. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. *Eng. Sci. Technol. Int. J.* 44:101455. doi: 10.1016/j.jestch.2023.101455

Zocholl, M., Stampouli, D., Wittfoth, M., and Mounier, G. (2025). Fundamental considerations for the use of explainable AI in law enforcement. *Front. Polit. Sci.* 7:1605619. doi: 10.3389/fpos.2025.1605619

Zowghi, D., and Rimini, D. (2023). 'Diversity and Inclusion in Artificial Intelligence. Boston, Massachusetts: Pearson Addison Wesley.

Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int. J. Hum. Rights* 24, 1572–1593. doi: 10.1080/13642987.2020.1743976