



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**AI Surveillance in Government: Governance, Accountability,
and Democratic Challenges in the Digital State**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της/του

Γεωργίου Ευστάθιου

Επιβλέπων: Χαράλαμπος Αλεξόπουλος, Επίκουρος Καθηγητής

Μέλη εξεταστικής επιτροπής: Επίκουρος Καθηγητής Χαράλαμπος Αλεξόπουλος, Επιβλέπων.
Καθηγητής Γιάννης Χαραλαμπίδης. Καθηγητής Ευριπίδης Λουκής

Σάμος, Σεπτέμβριος 2026

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πρόλογος και ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Επίκουρο Καθηγητή Χαράλαμπο Αλεξόπουλο, για την καθοδήγηση και την εμπιστοσύνη που μου έδειξε κατά τη διάρκεια της διπλωματικής μου εργασίας. Οι παρατηρήσεις του και η διαθεσιμότητά του βοήθησαν πολύ στο τελικό αποτέλεσμα.

Ευχαριστώ θερμά τη μητέρα μου και τον πατέρα μου, που με στήριζαν ηθικά και έμπρακτα σε όλα τα χρόνια των σπουδών μου και πίστεψαν σε μένα ακόμη και όταν εγώ δίσταζα. Οι κόποι τους για να σπουδάσω δεν ξεχνιούνται, και η ψυχραιμία τους με βοήθησε πολλές φορές να δω τα προβλήματα πιο απλά.

Ευχαριστώ επίσης τον αδερφό μου για την υπομονή του και για την υποστήριξη που μου έδινε όποτε τη χρειαζόμουν.

© 2026

του/της

Γεωργίου Ευστάθιου

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

Table of Contents

1	Introduction	1
2	Background	3
2.1	The Governance Gap in AI Surveillance: A Theoretical Framework.....	4
3	Methodology	6
3.1	Literature Search Strategy	6
3.2	Inclusion and Exclusion Criteria	7
3.3	Article Selection Process	8
3.4	Content Analysis Framework.....	8
3.5	Taxonomy Development.....	9
3.6	Limitations of the Study.....	10
4	Results: Table Analysis.....	12
4.1	Real-Time Facial Recognition (FRT)	15
4.1.1	Real-World Deployments & AI Techniques Employed	15
4.1.2	Accountability and Privacy Issues	15
4.1.3	Democratic, Ethical and Governance Concerns	15
4.2	Digital Contact Tracing Apps	16
4.2.1	Real-World Deployments & AI Techniques Employed	16
4.2.2	Accountability and Privacy Issues	17
4.2.3	Democratic, Ethical and Governance Concerns	17
4.3	Automated Gunshot Detection	17
4.3.1	Real-World Deployments & AI Techniques Employed	17
4.3.2	Accountability and Privacy Issues	18
4.3.3	Democratic, Ethical and Governance Concerns	18
4.4	Crowd Safety Manager	18
4.4.1	Real-World Deployments & AI Techniques Employed	19
4.4.2	Accountability and Privacy Issues	19
4.4.3	Democratic, Ethical and Governance Concerns	19
4.5	Autonomous Vehicle Sensor Fusion.....	20
4.5.1	Real-World Deployments & AI Techniques Employed	20
4.5.2	Accountability and Privacy Issues	21
4.5.3	Democratic, Ethical and Governance Concerns	21
4.6	Predictive Policing	21

4.6.1	Real-World Deployments & AI Techniques Employed	21
4.6.2	Accountability and Privacy Issues	22
4.6.3	Democratic, Ethical and Governance Concerns	22
4.7	Monitoring a Smart City	22
4.7.1	Real-World Deployments & AI Techniques Employed	22
4.7.2	Accountability and Privacy Issues	23
4.7.3	Democratic, Ethical and Governance Concerns	23
4.8	Udbetaling Danmark: AI in Welfare Surveillance	23
4.8.1	Real-World Deployments & AI Techniques Employed	23
4.8.2	Accountability and Privacy Issues	24
4.8.3	Democratic, Ethical and Governance Concerns	24
4.9	AI at Borders	24
4.9.1	Real-World Deployments & AI Techniques Employed	24
4.9.2	Accountability and Privacy Issues	25
4.9.3	Democratic, Ethical and Governance Concerns	25
4.10	AI Proctoring: The "Academicon"	26
4.10.1	Real-World Deployments & AI Techniques Employed	26
4.10.2	Accountability and Privacy Issues	26
4.10.3	Democratic, Ethical and Governance Concerns	26
4.11	AI-Driven Anti-Money Laundering Systems	27
4.11.1	Real-World Deployments & AI Techniques Employed	27
4.11.2	Accountability and Privacy Issues	27
4.11.3	Democratic, Ethical and Governance Concerns	27
5	Discussion	30
5.1	Interpretation of Findings	30
5.1.1	Real-World Deployments & AI Techniques Employed	31
5.1.2	Accountability and Privacy Issues	31
5.1.3	Democratic, Ethical and Governance Concerns	31
5.2	Common Democratic, Security and Ethical Considerations Across Solutions.....	32
5.2.1	Algorithmic Bias and Discriminatory Outcomes	32
5.2.2	Opacity and Black-Box Accountability	32
5.2.3	Erosion of Anonymity and the Right to Private Life	33
5.2.4	Normalisation of Surveillance Culture	33
5.2.5	Democratic Accountability and Public Contestation	33
5.2.6	Absence of Adequate Regulatory Frameworks	34
5.3	Emerging Risks in AI Surveillance	34
5.3.1	Generative AI, Synthetic Identities, and Deepfake	34
5.3.2	Adversarial Attacks and Cybersecurity Vulnerabilities	35
5.3.3	Interconnected Surveillance Ecosystems and Cross-Border Data Sharing.....	35

5.3.4	Autonomous Operation and the Limits of Human Oversight.....	35
6	Conclusion	37
	References	39

Table List

Table 1: I Surveillance Solutions in Government Operations 29

Figure List

Figure 1: Three-level taxonomy of AI surveillance solutions in government operations.	10
Figure 2: Real-world deployments & AI techniques employed (C1), accountability and privacy issues (C2), and democratic, ethical and governance concerns (C3) for Solutions 1–5.....	13
Figure 3: Real-world deployments & AI techniques employed (C1), accountability and privacy issues (C2), and democratic, ethical and governance concerns (C3) for Solutions 6–11.....	14

Περίληψη

Η παρούσα βιβλιογραφική ανασκόπηση εξετάζει τη σχέση ανάμεσα στις τεχνολογίες επιτήρησης με τεχνητή νοημοσύνη (TN) και τις κρατικές λειτουργίες, με έμφαση στην ιδιωτικότητα, τη λογοδοσία και τις δημοκρατικές προκλήσεις. Μέσω συστηματικής ανάλυσης της τρέχουσας ακαδημαϊκής βιβλιογραφίας, εντοπίστηκαν έντεκα μοναδικά εργαλεία επιτήρησης με TN, τα οποία οργανώθηκαν σε επτά τομείς: Δημόσια Υγεία, Επιβολή του Νόμου, Αστική Διαχείριση, Μεταφορές, Συνοριακό Έλεγχο, Εκπαίδευση και Χρηματοοικονομική Ρύθμιση. Κάθε εργαλείο αναλύθηκε βάσει τριών βασικών αξόνων: πραγματικές εφαρμογές και τεχνικές TN, ζητήματα λογοδοσίας και ιδιωτικότητας, και δημοκρατικές, ηθικές και θεσμικές ανησυχίες. Τα ευρήματα αναδεικνύουν ένα χάσμα διακυβέρνησης ανάμεσα στην τεχνική ικανότητα αυτών των συστημάτων και τη θεσμική ικανότητα εποπτείας τους, το οποίο εκδηλώνεται με συνεπείς τρόπους σε όλους τους τομείς: απώλεια ανωνυμίας σε δημόσιους χώρους, ευπάθεια συγκεντρωτικών δεδομένων, συνεχής παρακολούθηση πολιτών, αλγοριθμική μεροληψία, δυσανάλογες επιπτώσεις σε ευάλωτους πληθυσμούς και σταδιακή κανονικοποίηση της κουλτούρας επιτήρησης. Τα αποτελέσματα υποδεικνύουν άμεση ανάγκη για ρυθμιστικά πλαίσια που θα προστατεύουν τις δημοκρατικές αξίες και τα ανθρώπινα δικαιώματα εξίσου με την προώθηση της τεχνολογικής προόδου.

Το παρόν κεφάλαιο πρόκειται να δημοσιευθεί ως κεφάλαιο στον συλλογικό τόμο Handbook on Government Use of Artificial Intelligence (Edward Elgar Publishing, υπό έκδοση).

Λέξεις Κλειδιά: *Επιτήρηση με Τεχνητή Νοημοσύνη, αλγοριθμική διακυβέρνηση, χάσμα διακυβέρνησης, αλγοριθμική λογοδοσία, δημοκρατική εποπτεία, τεχνητή νοημοσύνη στον δημόσιο τομέα*

Abstract

This literature review examines how Artificial Intelligence (AI) Surveillance Technologies and Governmental Operations interact, specifically regarding Privacy and Ethics is conducted. Through a systematic analysis of current academic studies, 11 unique AI surveillance tools were identified and organized by sector, Public Health, Law Enforcement, Education, Transportation, Financial Regulation; each tool was analyzed using three primary factors: AI techniques used, privacy concerns, and ethical concerns. The results show that there are many common trade-offs between operational efficiency and fundamental rights. Typical privacy concerns include reduced anonymity in public areas, vulnerability of centralized data, and continuous citizen monitoring. The most typical ethics concerns include AI biases, disproportional impacts on marginalized populations, and normalization of surveillance cultures. The findings indicate an immediate need for regulatory systems that protect democratic values and human rights as much as they enable technological advancements.

This chapter is to be published as a chapter in the edited volume Handbook on Government Use of Artificial Intelligence (Edward Elgar Publishing, forthcoming).

Keywords: *AI surveillance, algorithmic governance, governance gap, algorithmic accountability, democratic oversight, public sector artificial intelligence*

1

Introduction

Artificial intelligence has dramatically changed how governments conduct surveillance around the world. At least 75 of the 176 countries listed in the Global Surveillance Index have committed to using and implementing artificial intelligence (AI) for their surveillance needs; especially in smart cities, facial recognition systems, and predictive policing [1]. Recent global events such as the COVID-19 pandemic have also contributed to this rapid growth, as biometric surveillance normalization was used to justify its use under the guise of protecting public health [2]. The introduction of artificial intelligence-based surveillance technologies into the way governments operate is a two-sided coin. On one hand, AI-based surveillance provides governments with new opportunities to enforce public safety, monitor public health, and enhance national security. On the other hand, AI-based surveillance raises important concerns regarding privacy violations, algorithmic discrimination, and the potential for an increase in authoritarianism [3]. As governments partner with large technology companies to implement increasingly advanced and widespread surveillance technologies, these concerns continue to grow [2]. This literature review provides a systematic examination of current AI-based surveillance solutions that are being implemented throughout various sectors of government. By reviewing extensive literature published within the last five years, we identified 11 unique AI-based surveillance applications that span from public health to law enforcement, education to transportation, border control to financial regulations. We focused our analysis on the specific AI techniques utilized in each application, the impact on individuals' privacy, and the ethical implications associated with each. We aim to provide policymakers, researchers, and civil society organizations with a framework for assessing the trade-offs between operational efficiency and protections of fundamental rights in the context of algorithmic governance.

The contribution of this chapter is threefold, and it may be useful to state at the outset what the chapter does and does not attempt. First, it provides a systematic cross-domain analysis of eleven AI surveillance solutions documented in real-world government deployment across seven governmental domains, examined along three consistent analytical axes. Second, and centrally, it develops a governance gap framework, set out in Section 2.1, which explains why the privacy, accountability, and democratic problems observed across these otherwise unrelated deployments are structurally reproduced rather than incidental to particular technologies or jurisdictions. Third, it derives from that framework a set of concrete governance recommendations addressed to the institutional conditions under which AI surveillance can be authorised, overseen, and held to account in democratic states. The chapter is therefore primarily a governance framework development exercise grounded in empirical case analysis, rather than a conceptual overview, a comparative regulatory study, or a technical assessment of surveillance systems, although it draws on international examples throughout and identifies systematic cross-jurisdictional comparison as a direction for future research.

2

Background

Forms of AI surveillance have proliferated in government functions, taking three major forms: smart city/safe city platforms, facial recognition, and “smart policing”. Smart city platforms combine sensors, cameras, and intelligent command centers to enable city management and public safety in real time. These systems have been installed in at least 56 countries around the world, such as Huawei’s Safe City programs in Serbia and Kenya, which use facial recognition and license plate recognition in hundreds of urban sites [1]. Facial recognition technology is one of the most controversial applications of AI, and enables biometric identification through database matching without the individual being identified having given their consent[4]. Approximately 64 countries have implemented facial recognition technology with this technology raising serious concerns about the potential for loss of individual privacy and the existence of bias within algorithms, and the fact that the technology disproportionately impacts communities of color [2]. In addition to its stationary installation, facial recognition technology is now available through the use of mobile body-worn cameras, as evidenced by the partnership between Malaysia and several Chinese technology companies to provide Malaysian police officers with facial recognition technology that can identify individuals in real-time. Predictive policing systems are designed to collect large amounts of data (geographic information, history of past arrests, biometric data, social media postings) to predict when and where crime may occur and to deploy police resources accordingly. However, like many other types of artificial intelligence, predictive policing systems can also perpetuate historical biases that are contained within the data that were utilized to train the algorithm, resulting in the increased presence of police in historically marginalized communities [1]. In addition to traditional law enforcement, AI surveillance is increasingly being employed in the areas of public health, border control, education, and financial regulation [2, 3].

As a result of the lack of comprehensive ethical frameworks and regulatory oversight, the widespread adoption of AI technologies during the COVID-19 pandemic has resulted in a need for systematic analysis of deployed solutions to develop policies based on empirical evidence, and to ensure that fundamental rights remain protected under the new paradigm of algorithmic governance.

2.1 The Governance Gap in AI Surveillance: A Theoretical Framework

The pattern observed across these deployments, technically capable systems whose oversight structures have not kept pace with their expanding use, is theorized in this chapter as a governance gap: the widening distance between what AI surveillance systems are technically capable of doing and the institutional capacity to authorize, constrain, and hold those systems accountable for what they actually do. This framework draws on five complementary bodies of theory.

Lyon's foundational work in surveillance studies describes how discrete, purpose-bound observation tends to evolve into diffuse surveillance "assemblages" that extend across public life without clear temporal or spatial limit [5]. This dynamic recurs throughout the case studies examined here, from facial recognition's shift from targeted investigation to routine biometric scanning, to the NYPD's Domain Awareness System, whose sensor network was justified as counterterrorism infrastructure before becoming a general-purpose policing platform. Zuboff's account of surveillance capitalism extends this observation to the logic of the institutions themselves: once data extraction is embedded within an infrastructure that citizens cannot easily opt out of, it becomes normalized as a matter of institutional routine rather than treated as an ongoing decision requiring justification [6]. AI proctoring and workplace monitoring illustrate this normalization directly, training students and employees to accept constant observation as an unremarkable condition of participation.

Danaher and colleagues' research agenda for algorithmic governance identifies the resulting legitimacy problem: when systems are deployed and iterated faster than the oversight mechanisms meant to govern them, a gap opens between a system's de facto authority over people's lives and the degree to which that authority has been democratically sanctioned [7]. Busuioc's analysis of algorithmic accountability specifies the mechanism by which this gap persists in practice: responsibility for an algorithmic system's failures is typically distributed across vendors, operating agencies, and the officials who act on a system's output, such that no single actor can be clearly held to account when the system errs [8]. This accountability deficit appears with particular clarity in the gunshot detection, crowd monitoring, predictive policing, and anti-money laundering cases, where independent audits, false alerts, and disputed flags routinely surface without a clear locus of institutional responsibility.

Finally, Moore's concept of public value and Trist's socio-technical systems theory provide the normative benchmark against which this governance gap is measured. Moore argues that public managers are accountable not merely for operational efficiency but for the broader public value their institutions create or erode [9]; a welfare fraud detection system, a smart city sensor network, or an AML pipeline may be technically efficient while still eroding the public

value, dignity, trust, and fair treatment, that its host institution exists to protect. Trist's socio-technical systems theory adds that technical and social subsystems must be jointly optimized for an institution to function well; deploying a technically sophisticated surveillance capability without a correspondingly developed social and institutional subsystem of consent, oversight, and redress produces exactly the imbalance this chapter documents across its eleven cases [10].

Together, these five perspectives motivate the analytical structure used throughout Chapter 4: each case study is examined along three consistent axes, Real-World Deployments & AI Techniques Employed, Accountability and Privacy Issues, and Democratic, Ethical and Governance Concerns, with the third axis operationalizing the governance gap framework directly. Where a case study documents function creep, opacity, or an unresolved accountability question, it is evidence of the same underlying gap between technical capability and institutional oversight that this framework predicts.

3

Methodology

To accomplish this literature review on the application of AI-based surveillance systems within government functions, a systematic process was used to find and assess the usage of AI-based surveillance systems within government functions. The method is an integration of systematic literature reviews methods, as well as taxonomy development methods, in order to perform a thorough assessment of operational implementations of AI-based surveillance systems. Overall, the approach followed the systematic literature review model developed by Bereton et al.[11]. The model provides a highly structured, and reproducible process for finding, assessing, and synthesizing relevant research.

3.1 Literature Search Strategy

A systematic literature review was performed through an examination of peer-reviewed articles from the past five years in the areas of computer science, social sciences, and interdisciplinary areas of interest in AI surveillance studies using the academic databases Scopus, Web of Science, and Google Scholar. The databases were chosen because of their comprehensive coverage of peer-reviewed literature related to AI surveillance studies. Peer reviewed articles published since 2019 were included to provide an overview of current AI surveillance technology and exclude outdated implementations which do not accurately represent current capabilities or privacy concerns associated with AI surveillance. The initial search utilized the keywords "AI Surveillance", "Facial Recognition", "Predictive Policing", "Smart City", and "Government Surveillance" in combination to identify peer-reviewed literature related to AI surveillance in government. Boolean operators (OR, AND) were used to limit or broaden the searches depending on the relevance.

Searches were also limited to "Government AND AI Surveillance", "Biometric Surveillance AND Facial Recognition OR Public Sector". A snowball technique was applied after the initial search for peer-reviewed literature utilizing the reference list of identified peer-reviewed literature to locate other sources of peer-reviewed literature not identified through the primary search strategy. This process continued iteratively until theoretically saturated, i.e., no additional peer-reviewed literature was found that meet the inclusion criteria.

3.2 Inclusion and Exclusion Criteria

The inclusion/exclusion criteria and systematic review protocol for selecting articles as outlined in [11] were clearly identified before the article selection process to assure objectivity and reproducibility. Journal articles and conference proceedings that had been peer reviewed were selected ahead of all other types of publications (gray literature, technical reports, etc.) since these provided an assurance of academic quality and credibility. In those instances where peer-reviewed references to governmental implementations lacked sufficient detail regarding technical aspects of the implementation, additional information from official government documentation or technical white-papers were added to provide a comprehensive representation of the surveillance solution.

Articles were included if they met the following criteria:

- Described AI surveillance technologies deployed in real-world government operations
- Provided sufficient technical detail regarding the AI techniques employed
- Discussed privacy implications or data handling practices
- Were published in English
- Focused on operational implementations rather than purely theoretical or conceptual frameworks
- Articles were excluded if they:
 - Discussed only proposed or hypothetical AI surveillance systems without real-world deployment
 - Lacked technical specificity regarding AI methods or privacy practices
 - Focused exclusively on private sector surveillance without government involvement
 - Were duplicate publications or pre-print versions later published in peer-reviewed venues

The emphasis on operational implementations, rather than conceptual articles, ensures that the analysis reflects actual privacy risks and ethical concerns arising from deployed systems, rather than theoretical possibilities.

3.3 Article Selection Process

The article selection process followed a multi-stage screening approach. In the initial screening phase, titles and abstracts of all retrieved articles were reviewed to eliminate obviously irrelevant sources. Articles that appeared potentially relevant proceeded to the full-text review stage, where they were assessed against the inclusion and exclusion criteria described above. During this stage, articles were evaluated for the depth of technical detail provided, the clarity of privacy and ethical discussions, and evidence of real-world deployment. A total of 42 articles were initially identified searching the database and snowballing. After removing duplicates, 31 unique articles remained. Following the selection of the title and abstract screening, 21 articles proceeded to full-text review. Of these, 11 articles met all inclusion criteria and formed the basis for the final analysis. The most common reasons for exclusion during full-text review were lack of operational deployment evidence and insufficient technical detail regarding AI implementation.

3.4 Content Analysis Framework

The content analysis methodology developed by Saheb [2] was adapted and applied to extract relevant information from the selected articles. This approach involves systematic coding of textual data to identify patterns, themes, and key characteristics of AI surveillance implementations. Each article was read in full, and relevant passages discussing technical implementations, privacy practices, and ethical concerns were highlighted and coded. For each identified AI surveillance solution, information was extracted and organized according to three primary analytical dimensions, mirroring the chapter title's central concerns with governance, accountability, and democratic challenges: (1) **Real-World Deployments & AI Techniques Employed:** This dimension captures the specific artificial intelligence and computational methods employed by the surveillance system, alongside documented evidence of its real-world deployment by government or institutional actors. Categories include computer vision algorithms (such as convolutional neural networks for image recognition), machine learning classifiers (such as support vector machines or random forests), natural language processing techniques, deep learning architectures, and sensor fusion approaches. The goal was to understand not only what the system does but how it accomplishes its surveillance objectives from a technical standpoint, and where it has actually been deployed. (2) **Accountability and Privacy Issues:** This dimension examines how personal data is collected, stored, processed, and shared within the surveillance system, together with the accountability mechanisms, or their absence, that govern who may be held responsible when the system errs. Specific attention was paid to whether informed consent is required or obtained, the scope and granularity of data collection (e.g., real-time video feeds, biometric identifiers, location data), data retention policies, access controls, whether data is shared with third parties or across government agencies, and whether a clear locus of institutional responsibility exists for the system's outputs. This analysis helps identify the specific privacy and accountability risks posed by each surveillance solution. (3) **Democratic, Ethical and Governance Concerns:** This dimension addresses broader ethical and democratic questions beyond privacy, including algorithmic bias and discrimination, fairness and equity in application,

impacts on democratic freedoms and civil liberties, public contestation and resistance, and potential for mission creep or function expansion beyond the system's original purpose. Ethical analysis was based on frameworks from critical surveillance studies, algorithmic fairness research, and human rights literature.

3.5 *Taxonomy Development*

The iterative development of the final taxonomy followed the iterative methodology of Nickerson et al. [12] who describe an organized approach to building taxonomies in IS research using the methods of empirical- to-conceptual and conceptual-to-empirical to iteratively refine the dimensions and characteristics of the taxonomy as determined through both theoretical foundations and empirical evidence from the collected data. The taxonomy development process involved the following stages: **Stage 1 - Determination of meta-characteristics:** The objective of developing the taxonomy was defined as categorizing AI surveillance solutions by their application domains in government, technology characteristics and privacy-ethics implications. **Stage 2 - Specification of ending conditions:** The taxonomy development process would cease to continue when: (a) All identified surveillance solutions were categorized into the developed taxonomy, (b) At the highest level, each category of the taxonomy was mutually exclusive and (c) The taxonomy provided meaningful distinctions to compare privacy and ethics implications among the various implementations of the solutions. **Stage 3 - Iterative identification of dimension and characteristic:** Through iterative analysis of the selected articles, seven primary domains of government emerged as natural categories to organize AI surveillance solutions: (1) Public Health, (2) Law Enforcement, (3) Urban Management, (4) Transportation, (5) Border Control, (6) Education, and (7) Financial Regulation. The categories were iteratively refined to test whether each identified surveillance solution could be meaningfully classified and whether the categories captured the necessary differences in application context. **Stage 4 - Validation and Refining:** Each surveillance solution was independently classified by validating that it belonged to one primary domain category. Those surveillance solutions that appeared to belong to multiple domains were analyzed to determine its primary purpose and context of use. Those solutions that were implemented in a real world environment and had sufficient detail about their technical implementation, privacy practices, and ethical implications were included in the final taxonomy. This decision rule insures that the analysis represents deployed systems that have documented impacts rather than hypothetical or poorly documented cases. The systematic evaluation produced a taxonomy of 11 different applications of AI surveillance across the seven governmental domains. These solutions are presented in table format in Section 4 to enable comparisons of the real-world deployments and AI techniques employed, the accountability and privacy issues raised, and the democratic, ethical and governance concerns associated with the various surveillance contexts. The systematic nature of this methodology, following established guidelines for both systematic literature reviews [11] and taxonomy development [12], enables the identification of repeated privacy and ethics issues across the individual domains and the discovery of patterns in

how AI surveillance is used and the common challenges it poses to individual rights and democratic governance.

The resulting taxonomy is illustrated in Figure 1, presenting a three-level hierarchical structure: the root node (AI Surveillance), seven governmental sectors, and eleven deployed solutions.

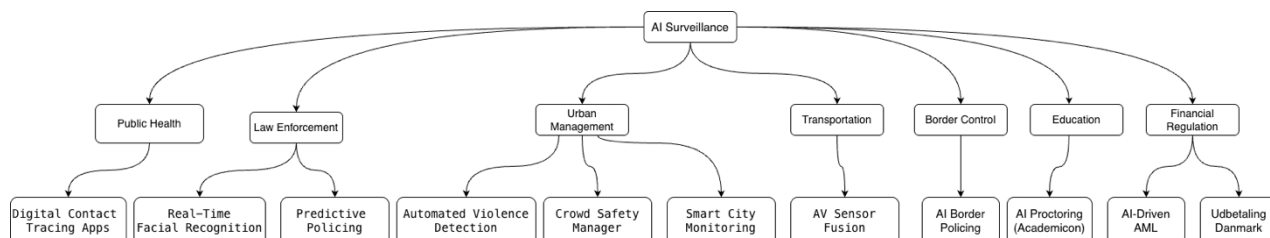


Figure 1: Three-level taxonomy of AI surveillance solutions in government operations.

3.6 Limitations of the Study

Four limitations should be borne in mind when interpreting the findings presented in this chapter. First, the systematic literature review was restricted to peer-reviewed sources published in English since 2019. This constraint improves comparability and currency, but it necessarily excludes relevant grey literature, national-language policy documentation, and implementations documented primarily outside Anglophone scholarship, meaning that the eleven cases analysed here are likely to over-represent European and North American deployments relative to their global distribution.

Second, the analysis relies on publicly available documentation. Surveillance systems that are classified, subject to commercial confidentiality, or deployed without public disclosure are by definition inaccessible to academic review. The governance gap documented in this chapter is therefore measured against the systems whose existence and operation have been made public, and the true gap may be wider than what the available evidence permits us to demonstrate.

Third, the content analysis framework described in Section 3.4 involves interpretive judgement in coding passages against the three analytical axes. The coding was carried out by the authors without a formal interrater reliability assessment, and different analysts might reasonably assign some material differently, particularly at the boundary between the accountability and privacy axis and the democratic, ethical and governance axis, where the two concerns frequently overlap.

Fourth, the governance gap framework advanced here is conceptual rather than empirically tested. It is derived from and illustrated by the eleven cases examined, but it has not been validated against a representative sample of jurisdictions or subjected to quantitative measurement. Establishing whether the gap varies systematically across regulatory regimes, and identifying which institutional arrangements narrow it most effectively, remains an open empirical question that this chapter identifies rather than resolves.

4

Results: Table Analysis

Our systematic analysis identified 11 distinct AI surveillance solutions currently deployed across government operations (1). These various types of surveillance systems fall into a number of categories: Public Health (Digital Contact Tracing Apps), Law Enforcement (Real-Time Facial Recognition, Automated Gunshot Detection, Predictive Policing), Urban Management (Crowd Safety Manager, Monitoring a Smart City), Transportation (Autonomous Vehicle Sensor Fusion), Border Control (AI at Borders), Education (AI Proctoring), and Financial Regulation (Udbetaling Danmark, AI-Driven Anti-Money Laundering Systems).

Figures 2 and 3 provide a structured overview of each solution across the three analytical dimensions examined in this review: real-world deployments & AI techniques employed (C1), accountability and privacy issues (C2), and democratic, ethical and governance concerns (C3).



Figure 2: Real-world deployments & AI techniques employed (C1), accountability and privacy issues (C2), and democratic, ethical and governance concerns (C3) for Solutions 1–5.



Figure 3: Real-world deployments & AI techniques employed (C1), accountability and privacy issues (C2), and democratic, ethical and governance concerns (C3) for Solutions 6–11.

4.1 Real-Time Facial Recognition (FRT)

Real-time facial recognition is one of the most intrusive forms of AI-based government surveillance and is also one of the most widespread. As of 2020, at least 64 countries were utilizing facial recognition technology in their surveillance infrastructure [1].

4.1.1 Real-World Deployments & AI Techniques Employed

Real-time facial recognition works by using CCTV cameras to capture facial images, identifying facial characteristics and generating biometric templates, and then comparing those against a vast database of stored images, all done without the consent of the individual being tracked [4]. Deep learning algorithms based on convolutional neural networks (CNNs) are used to identify and analyze facial features, enabling the system to operate across various lighting conditions and camera angles; biometric matching algorithms then compare the resulting facial feature data to generate a similarity score. In practice, this technical capability has translated into wide-ranging deployments. Between 2016 and 2019, the UK Metropolitan Police conducted ten trials of the technology and found an 81% error rate, raising concerns about accuracy

[1]. In 2016, Germany installed 26 stationary cameras with facial recognition capabilities at the main station in Cologne, while Huawei's Safe City project in Belgrade plans to install over 800 such cameras [1]. Malaysian police are currently equipping officers with facial recognition cameras that allow real-time comparison of facial data against a database.

4.1.2 Accountability and Privacy Issues

FRT removes anonymity in public space by creating permanent biometric records of an individual's movements and relationships. It operates continuously and without time constraints, collecting facial data of every person who enters a surveilled area regardless of whether they are suspected of any wrongdoing [4]. Access to the resulting databases is largely unregulated: many U.S. law enforcement agencies scan millions of driver's license photographs without either the knowledge or the consent of the individuals photographed [1]. This continuous, consent-free accumulation of biometric records exemplifies what Lyon identifies as the shift from bounded, purpose-specific surveillance relations toward diffuse "assemblages" that extend across public life with no clear temporal or spatial limit [5], and the resulting lack of any single body responsible for auditing how these databases are used or shared leaves affected individuals with no clear channel through which to seek accountability.

4.1.3 Democratic, Ethical and Governance Concerns

There is considerable evidence of race and gender disparities in facial recognition systems. Studies have found false positive rates ranging from 1 to 35 percent for dark-skinned females, compared to 0.2 percent for white males in optimal testing environments [1, 3]. Travelers with darker skin tones experience false rejection rates of over 20 percent at some European border checkpoints,

leading to detentions and denial of asylum claims [3]. These disparities stem from biased training datasets that underrepresent minority groups, illustrating how automated decision-making systems can reinforce systemic inequality [2].

Beyond individual bias, FRT raises broader questions of democratic accountability. Cities such as San Francisco, Oakland, and Somerville have passed ordinances banning government use of facial recognition on these grounds, while during the Hong Kong protests, protesters disabled facial recognition on their phones to avoid identification by law enforcement [1]. The absence of universally applicable human rights frameworks for facial recognition allows governments and private companies to unilaterally expand its use, increasing the vulnerability of specific populations, including minorities, children, and marginalized groups [2]. This pattern is consistent with Zuboff's account of how surveillance infrastructures, once normalized under the banner of public safety, extract behavioral data as a matter of routine and place the burden of resistance on individuals rather than on the institutions deploying the technology [6]. It also illustrates the accountability deficit that Busuioc identifies as structural to algorithmic systems: because responsibility for FRT errors is distributed across vendors, police departments, and municipal authorities, no single actor can be clearly held to account when the system misidentifies or wrongfully flags an individual [8]. Read through Danaher and colleagues' research agenda for algorithmic governance, the FRT case shows precisely the kind of legitimacy gap that arises when effective deployment outpaces the oversight structures meant to constrain it [7].

4.2 Digital Contact Tracing Apps

The COVID-19 pandemic created a new demand for digital contact tracing applications, which automate many of the steps involved in traditional human-intensive contact tracing and promise faster, more widespread tracking than human contact tracing teams [4].

4.2.1 Real-World Deployments & AI Techniques Employed

Contact tracing applications rely on wireless technologies to detect nearby devices and exchange "digital keys" to document encounters; once a user reports a positive test result, all applications that stored their unique identifier send a notification to that user's contacts [4]. Implementations differ along two main technical axes: the database architecture used to store user identifiers (centralized versus decentralized), and the technology used to measure proximity (GPS, Bluetooth Low Energy, WiFi MAC address scanning, or cellular geolocation) [4]. Singapore's TraceTogether, introduced in March 2020, used short-range Bluetooth proximity sensing and retained data for roughly 21 days, while Germany's Corona-Warn-App stored all data locally on each user's phone rather than on centralized government servers and required no personally identifiable information [4].

4.2.2 Accountability and Privacy Issues

Centralized architectures transmit users' contact lists to central platforms so that authorities can determine whether contact occurred with an infected user, a design that creates substantial privacy risk because it moves sensitive data out of users' direct control [4]. Decentralized approaches instead store data locally on users' phones and merely perform pull and push operations against lists of positive-tested keys; Bluetooth proximity detection, which measures signal strength rather than absolute location, is consequently less invasive than GPS-based tracking [4]. Norway's Smittestop initially relied on real-time GPS transmission to central servers, drew widespread criticism as an example of mass surveillance, and was ultimately withdrawn and revised [4]. China's Alipay Health Code combined GPS and self-reported data to generate infection risk scores and required mandatory adoption, restricting users' freedom of movement based on an opaque, centrally calculated score [4].

4.2.3 Democratic, Ethical and Governance Concerns

Implementation was dominated by a voluntary versus mandatory adoption debate: while most countries encouraged voluntary installation, compulsory requirements in certain jurisdictions raised concerns about state coercion [2]. The effectiveness of digital contact tracing depended heavily on user acceptance, which research has linked to at least eight distinct factors, including trust, privacy concern, technological literacy, and users' willingness to trade privacy for reduced health risk [4]. The gap between the most privacy-protective implementations (TraceTogether, Corona-Warn-App) and the most invasive ones (Smittestop, Alipay Health Code) illustrates a recurring pattern in this analysis: absent a clear regulatory framework in place before deployment, rapidly built emergency technologies can just as easily entrench mass surveillance as protect public health, with the difference resting less on the underlying AI technique than on the governance choices made around it.

4.3 Automated Gunshot Detection

Acoustic gunshot detection technology (GDT), most prominently the ShotSpotter system (now marketed by SoundThinking), is a network of acoustic sensors deployed across urban areas to identify gunfire and alert police in real time. As of the early 2020s, GDT was in use in roughly ninety American cities [13].

4.3.1 Real-World Deployments & AI Techniques Employed

GDT relies on networks of twenty to twenty-five acoustic sensors per square mile, typically mounted on rooftops and utility poles, which listen for impulsive sounds consistent with gunfire and triangulate their point of origin [13]. When a sensor array registers a probable gunshot, the recording is forwarded to a monitoring center where human reviewers, often former law enforcement personnel, confirm whether the sound is gunfire before an alert is dispatched to police, typically within thirty to forty-five seconds of the triggering sound [13]. Some deployments integrate GDT with CCTV networks, allowing cameras to reorient automatically toward the location of a detected alert [13]. Empirical assessments of the technology's accuracy

vary considerably by jurisdiction and methodology: reported detection rates for genuine gunfire have ranged from roughly 70 percent in St. Louis to over 99 percent in Charleston, while independent audits, such as an investigation of the Chicago Police Department's use of the system, found that only 9 percent of alerts corresponded to confirmed evidence of a gun-related offense [13].

4.3.2 *Accountability and Privacy Issues*

GDT sensors continuously monitor ambient sound in the areas where they are deployed, and because they begin recording briefly before and after a triggering sound, they have in some documented cases picked up fragments of human speech and social interaction near the sensor, raising concerns about incidental audio surveillance beyond the system's stated gunshot detection purpose [13]. Sensor placement itself raises further privacy and equity concerns: because GDT networks are concentrated in areas that police departments designate as high crime, their surveillance footprint disproportionately covers Black and Latino neighborhoods, effectively subjecting residents of those areas to a persistent, involuntary level of acoustic monitoring that residents of other neighborhoods do not experience [14].

4.3.3 *Democratic, Ethical and Governance Concerns*

Independent evaluations of GDT deployments have consistently found that alerts rarely correspond to confirmed gun crimes and that the technology's accuracy claims, largely generated by the vendor itself, are difficult to verify independently [13]. Longitudinal analysis across large metropolitan counties similarly found no consistent evidence that GDT deployment reduces firearm homicides, despite its considerable cost to police departments [15]. Because alerts concentrate police attention and stop and frisk activity in the same historically over-policed neighborhoods where sensors are installed, the technology has been shown to reproduce, rather than correct, existing racial disparities in enforcement outcomes [14]. This pattern illustrates the kind of accountability deficit that Busuioc identifies as structural to algorithmic and sensor-driven policing tools: responsibility for a false alert, and for the police response it triggers, is distributed across the vendor, the monitoring center, and the responding officers, such that no single actor can be clearly held to account when an alert proves unfounded [8]. Legal scholarship has further argued that GDT-generated alerts are increasingly used by courts and police as a substitute for the individualized suspicion ordinarily required before a stop, raising constitutional concerns beyond the immediate accuracy debate [13].

4.4 *Crowd Safety Manager*

Crowd safety management utilizes AI-based crowd counting and group detection to monitor crowds at public gatherings, large events and urban spaces. A prominent real-world example of this class of system is the New York City Police Department's Domain Awareness System (DAS), a citywide network of sensors, cameras, and databases that delivers real-time analytics on public spaces to officers' smartphones and precinct desktops [16].

4.4.1 Real-World Deployments & AI Techniques Employed

DAS links together roughly 9,000 CCTV cameras and 500 license plate readers, combined with radiation and chemical sensors, gunshot detection feeds, and machine learning tools such as Patternizr, which connects unsolved crimes across the city [16, 17]. Computer vision algorithms process the resulting video streams to count individuals in public areas, identify group formations, and detect abnormal crowd patterns such as rapid dispersion or dense congestion that may indicate a safety concern [16]. The system was developed jointly by the NYPD and Microsoft beginning in 2008 and has since been credited with generating an estimated \$50 million per year in operational savings for the city [16].

4.4.2 Accountability and Privacy Issues

DAS was originally authorized under the NYPD's Public Security Privacy Guidelines as a tool to "monitor public areas and public activities" for counterterrorism purposes [17]. In practice, its scope has expanded well beyond that original mandate: DAS analytics are now used to inform routine patrol decisions, identify loiterers, and support general criminal investigations, none of which were contemplated in the system's founding privacy policy [17]. Because the same privacy guidelines govern both the original counterterrorism use and this expanded general-policing use, DAS operates with weaker oversight than a purpose-built crowd or criminal database would require, and the department has stated that it does not archive the predictions its analytics generate, which forecloses independent study of the system's outputs over time [17].

4.4.3 Democratic, Ethical and Governance Concerns

The migration of DAS from a narrow counterterrorism tool to a general-purpose policing platform is a documented instance of the function creep that crowd monitoring systems are broadly susceptible to: sensors and cameras justified by a specific, bounded purpose become, over time, the infrastructure for much broader forms of social observation [17]. This expansion has occurred with minimal external accountability. Much of what is publicly known about DAS's inputs and outputs comes not from proactive disclosure but from a three-year public records dispute between the Brennan Center and the NYPD, and the department has resisted extending its existing oversight policies to cover newer analytics, such as predictive policing tools, that are technically housed within the same system [17]. This pattern illustrates the accountability deficit that Busuioc identifies as characteristic of algorithmic public sector tools more broadly: without archived outputs or a public policy that keeps pace with the system's expanding capabilities, there is no reliable mechanism by which affected residents, or even city oversight bodies, can determine whether the system's crowd and behavior analytics are being applied fairly across neighborhoods [8].

4.5 Autonomous Vehicle Sensor Fusion

Traffic surveillance camera systems have become an important source of external data for autonomous vehicle (AV) applications, and the fleets of AVs themselves have, in turn, become a source of surveillance data for government agencies. This bidirectional relationship illustrates how transportation infrastructure has begun to overlap with public surveillance networks.

4.5.1 Real-World Deployments & AI Techniques Employed

Autonomous vehicles traditionally rely on onboard sensor suites, including LiDAR, RADAR, cameras, GPS, and inertial measurement units, to sense their surroundings and navigate safely; advanced implementations additionally use municipal traffic surveillance cameras as both an extra image source and an anchor node for AI networking systems, allowing four-dimensional detection, precise localization, and real-time positioning that exceeds what onboard sensors alone can achieve [18]. This fused architecture allows AVs to perceive conditions outside their immediate sensor range, for example around corners, through intersections, or in tunnels where GPS signals are unreliable [18]. Separately from this technical fusion with municipal infrastructure, each AV independently generates a continuous, high-resolution record of its surroundings: a single vehicle can carry more than a dozen cameras providing 360-degree coverage, alongside LiDAR and radar [19].

4.5.2 Accountability and Privacy Issues

Because AVs record their surroundings continuously as a byproduct of navigation, they generate a mobile surveillance record of public space that is retained by the operating company, typically for weeks to months, and can be requested by police [19]. A 2023 investigation found that San Francisco police had, per their own training materials, used autonomous vehicle footage for investigations on multiple occasions, and that police in San Francisco and Maricopa County, Arizona had served at least nine warrants to Waymo and Cruise for vehicle footage between 2021 and mid-2023, covering cases ranging from hit-and-runs to homicides [19]. Data retention and disclosure practices vary considerably between vehicle operators, with no consistent industry standard governing how long footage is kept or under what conditions it is shared with law enforcement [19].

4.5.3 Democratic, Ethical and Governance Concerns

The same infrastructure that improves AV safety functions, unlike a single stationary CCTV camera, as a continuously moving sensor platform that can create dynamic, city-wide maps of where specific vehicles and individuals travel [19]. Digital rights researchers have noted that, unlike purpose-built law enforcement surveillance tools, AV data access typically proceeds without the transparency requirements or public debate that would normally accompany the creation of a new municipal surveillance capability [19]. Because this capability emerged as a byproduct of a distinct commercial technology, oversight has lagged the practice itself: police departments have begun treating AV fleets as a source of investigative footage before legislatures or courts have clarified what legal standard, if any, should govern such requests, echoing the pattern seen with acoustic gunshot detection technology, where legal accountability struggled to keep pace with an already-normalized law enforcement practice [13].

4.6 Predictive Policing

Predictive policing applies machine learning to predict potential locations for crime and to identify individuals statistically likely to be involved in future crime. Systems such as PredPol use historical crime data, arrest records, and geographical patterns to inform officer deployment; at least sixty U.S. police departments have employed systems of this kind, while Chinese authorities have incorporated predictive policing into broader surveillance networks, most notably in Xinjiang, where it has been used alongside systematic targeting of ethnic minorities [1].

4.6.1 Real-World Deployments & AI Techniques Employed

Predictive policing systems use natural language processing, pattern recognition, and data mining to analyze large historical crime datasets, correlating factors such as time of day, geographic location, offense type, and environmental conditions to produce probability assessments of future crime [1]. As new crime reports and arrests are added, the models continually update. A widely cited demonstration applied a published PredPol-style algorithm to drug crime records in Oakland and found that the model would have concentrated policing overwhelmingly in low-income,

minority neighborhoods, regardless of where drug use was estimated to actually occur across the city's population [20].

4.6.2 Accountability and Privacy Issues

Because predictive systems draw on historical arrest data, they aggregate detailed records of prior police contact, demographic characteristics, and geographic patterns into risk scores applied to individuals who have not themselves been suspected of any specific offense [20]. This represents a significant departure from traditional, suspicion-based policing: individuals are subjected to heightened scrutiny based on correlational patterns in aggregated data about the areas or social groups to which they belong, rather than on evidence connected to their own conduct [1].

4.6.3 Democratic, Ethical and Governance Concerns

Because predictive models are trained on historical arrest data that reflects decades of biased policing practices, they tend to replicate and amplify those biases: neighborhoods with heavier historical police presence generate more arrest data, which predictive systems interpret as evidence of higher crime, which in turn justifies further concentrated policing [20]. Formal modeling of this dynamic confirms that predictive policing systems are susceptible to a runaway feedback loop, in which police are repeatedly directed back to the same neighborhoods regardless of the true underlying crime rate, since the system observes only the crime that its own deployment decisions cause it to discover [21]. Because of the proprietary nature of most predictive policing algorithms, affected communities and even police departments themselves often cannot determine how predictions are generated or why particular neighborhoods are targeted, producing the same accountability deficit that characterizes algorithmic public-sector tools more broadly: when a prediction proves wrong, the algorithm vendor, the police department, and the data providers can each point to the others as responsible [8]. Whether predictive policing meaningfully reduces crime, as opposed to simply directing more police attention and therefore more arrests toward the same neighborhoods, remains empirically contested [20].

4.7 Monitoring a Smart City

Smart cities deploy extensive IoT sensor, camera, and data collection systems throughout urban infrastructure, ostensibly to optimize municipal operations and improve service delivery. The same infrastructure also creates an extensive surveillance system capable of monitoring citizens' behavior and movement.

4.7.1 Real-World Deployments & AI Techniques Employed

Sensors integrated into streetlights, traffic signals, public transportation, waste management, and environmental monitoring equipment continuously collect data on traffic flows, pedestrian movement, air quality, and use of public spaces, which AI algorithms process to generate patterns and inform municipal decisions in real time [1]. Technology vendors market these capabilities explicitly as security infrastructure: Huawei's "Safe City" platforms combine video surveillance, facial recognition, incident command systems, and big data analysis to support law enforcement,

and have been deployed in at least 56 countries, including a planned installation of over 800 camera locations with facial recognition capability in Belgrade [1].

4.7.2 Accountability and Privacy Issues

Smart city sensors collect detailed information about residents' daily habits, movements, associations, and use of public services; when combined and analyzed together, this data can produce a highly detailed account of an individual's life, including where and when they move and whom they interact with [1]. Unlike historical, targeted forms of surveillance, smart city monitoring captures data on all members of society continuously and by default, creating permanent records of urban activity rather than records generated by a specific investigation [2].

4.7.3 Democratic, Ethical and Governance Concerns

Infrastructure originally justified by narrow, benign functions, such as managing traffic flow or optimizing waste collection, is commonly repurposed for more invasive surveillance uses, including tracking protest activity or profiling populations, typically without public knowledge or consent [2]. This blurring of "urban management" and "state control" allows comprehensive monitoring to expand with minimal public opposition, since authorities frame the underlying infrastructure as a service improvement rather than a surveillance capability [2]. Smart city surveillance infrastructure is typically procured and deployed through public-private partnerships that occur largely outside of democratic scrutiny, with limited public consultation or privacy impact review, leaving residents unable to determine what data is collected, how it is analyzed, or who has access to it [1]. This pattern mirrors the trajectory already observed in the NYPD's Domain Awareness System, where infrastructure justified by one narrow purpose was, over time, redirected toward much broader forms of population monitoring with comparatively little external accountability [17].

4.8 Udbetaling Danmark: AI in Welfare Surveillance

Udbetaling Danmark (UDK) is a centralised Danish public authority responsible for administering a broad range of welfare benefits, including pensions, child allowances, housing support, and unemployment benefits. In 2018, UDK transferred approximately 222 billion DKK in social benefits to approximately 2.4 million persons [22]. To manage this scale of distribution, UDK established a cross-cutting data unit in 2014 to develop data-driven profiling models aimed at detecting fraudulent or erroneous benefit payments.

4.8.1 Real-World Deployments & AI Techniques Employed

By 2019, UDK employed approximately 60 different predictive models to analyse and combine data from multiple public registers, including the central personal register (CPR), the income register, tax records, regional health data, company registers, and increasingly, publicly available data from social network sites [22]. The models range from simple decision trees that classify individuals as high or low risk to more complex machine learning models. On a regular basis, the

system scans the full dataset for a specific benefit type and generates a ranked list of individuals statistically flagged as being at high risk of receiving benefits to which they are not entitled.

4.8.2 Accountability and Privacy Issues

The UDK system exemplifies what Jørgensen terms "ubiquitous dataveillance": the continuous, automated monitoring of an entire population of benefit recipients without their knowledge or prior consent [22]. The profiling occurs "behind the back" of the individual, with no notification that their data across multiple state systems is being cross-referenced and scored. In a welfare state such as Denmark, where the majority of families receive some form of public benefit, the scope of this surveillance is effectively population-wide, and the models also process data relating to individuals' family members and social connections, extending surveillance beyond the direct subject of investigation [22].

4.8.3 Democratic, Ethical and Governance Concerns

The historical data used to train UDK's models inherently reflects prior enforcement priorities and reporting patterns, meaning that groups historically subject to greater scrutiny, such as low-income households or immigrant communities, are more likely to be flagged by the system; smaller benefit datasets also produce models with higher error rates, increasing the risk of false positives for recipients of less common benefits [22]. This creates a structural asymmetry in which the system disproportionately targets vulnerable citizens who depend on welfare support, while other forms of financial non-compliance, such as tax evasion by wealthier individuals, remain largely outside its scope. Automated profiling also makes it significantly harder for citizens to understand why they have been flagged or to appeal automated determinations, effectively rendering the citizen a collection of data points rather than a rights-bearing individual [22]. Jørgensen argues that, absent a more human-centric approach to algorithmic governance, the digital welfare state risks becoming a form of digital technocracy that treats citizens as objects of calculation and prediction rather than as subjects with agency, dignity, and enforceable rights [22], a concern that resonates with Moore's account of public value: a welfare system that optimizes for fraud detection at the expense of trust and dignity may be technically efficient while failing the broader public value it exists to serve [9].

4.9 AI at Borders

The use of artificial intelligence has significantly changed the way European border control polices borders, with predictive analytics, biometrics, and risk assessment systems increasingly used to track migrants and asylum seekers [3]. Border agencies process large amounts of data, including satellite imagery, sensor feeds, and past migration information, to predict where and when unauthorized crossings are likely to occur and to direct personnel accordingly [3].

4.9.1 Real-World Deployments & AI Techniques Employed

The European Border Surveillance System (EUROSUR) links together multiple data sources to provide real-time risk assessments to border authorities [3]. Biometric identification methods,

including facial recognition, fingerprints, and eye scans, are used to protect the EU's external borders; the Schengen Information System (SIS II) and the Entry/Exit System (EES) link biometric information together to track travelers and identify unauthorized entries [3]. Risk-assessment systems such as Advanced Passenger Information (API) and Passenger Name Record (PNR) classify travelers based on predefined profiles examining travel history, behavioral patterns, and socio-economic factors [3]. The EU has also experimented with automated lie-detection technologies such as iBorderCtrl, piloted in Greece, Hungary, and Latvia, which claimed to assess traveler honesty from facial micro-expressions during questioning [3].

4.9.2 Accountability and Privacy Issues

Databases containing visa applications, asylum claims, passenger records, and criminal records are increasingly interconnected via algorithms with little transparency, allowing border authorities to track individuals' activity across multiple systems and jurisdictions and to build detailed dossiers of their interactions with border control [3]. There is currently no comprehensive legal framework governing the sharing of this personal data among EU member states, which risks establishing a de facto regime for controlling human mobility within the EU without corresponding human rights safeguards [3].

4.9.3 Democratic, Ethical and Governance Concerns

Predictive border models do not represent objective security risk so much as the priorities of political leaders, generating broad alerts that justify enforcement action without allowing for humanitarian assessment; because these models rely on historical migration data, they reinforce existing biases and direct disproportionate attention toward specific nationalities and ethnic groups [3]. Research demonstrates significant bias in biometric identification specifically affecting darker-skinned individuals: false rejection rates for African and Middle Eastern travelers have exceeded 20 percent at several checkpoints, leading to wrongful detention or denial of asylum claims, illustrating how ostensibly neutral biometric systems can produce racial hierarchies at the heart of border enforcement [3]. Psychology experts have widely rejected the "affect recognition science" underlying tools such as iBorderCtrl as scientifically unproven, yet authorities continued to explore such pseudoscientific methods despite the evidence against their validity [3]. The opacity of the algorithms underlying these risk-assessment systems makes it difficult to challenge incorrect classifications, and because decisions are generated automatically, liability cannot be clearly assigned to algorithm developers, border authorities, or the private companies operating the systems, undermining both the EU legal principle of accountability and the right to an effective remedy [3]. Taken together, these systems represent what can be described as a form of algorithmic sovereignty, in which machine-based risk assessment increasingly governs human mobility, with accountability distributed among public agencies and private technology vendors in ways that make no single actor answerable for the system's failures [3, 8].

4.10 AI Proctoring: The "Academicon"

AI-based proctoring monitors students during online assessments via webcam, screen capture, keystrokes, and biometrics, illustrating what Swartz and McElroy describe as the emergence of an "academicon", a higher education environment structured around algorithmic oversight [23].

4.10.1 Real-World Deployments & AI Techniques Employed

Proctoring systems such as Proctorio combine several surveillance modalities simultaneously: webcam monitoring tracks facial expressions, eye movements, and head movements to flag "suspicious" behavior; biometric comparison matches a student's facial characteristics against a stored reference image to confirm identity; screen capture logs every on-screen action; browser lockdown restricts access to unauthorized resources; and algorithmic analysis of keystroke dynamics and mouse patterns detects deviations from a machine-learned baseline of "normal" exam behavior [23].

4.10.2 Accountability and Privacy Issues

Proctoring systems require students to grant extraordinary access to their private spaces and personal devices: webcams record students' home environments, including who else is present and what personal items are visible, creating an archive of domestic life that extends well beyond what is necessary to assess academic honesty [23]. This transforms students' homes into monitored test-taking locations, collapsing the boundary between educational and personal space, and many systems retain these recordings indefinitely [23]. AI-based proctoring is also only one part of a wider academic surveillance infrastructure: plagiarism-detection software retains copies of student submissions indefinitely, learning management systems log login times and page views, and predictive models generate "risk scores" estimating which students are likely to underperform, together producing comprehensive profiles of students' intellectual habits and behavior [23].

4.10.3 Democratic, Ethical and Governance Concerns

AI-based proctoring systems demonstrate significant racial and disability bias: facial recognition components commonly fail to recognize dark-skinned students, costing them valuable test-taking time, while students with physical disabilities, neurodivergent students, and students with anxiety disorders receive disproportionate false positives for academic dishonesty when stress responses such as fidgeting or looking away are misclassified as cheating [23]. Consent in this context is largely nominal rather than genuine: students face an effective choice between accepting invasive surveillance or forgoing their education, an imbalance of institutional power that precludes meaningful voluntariness, particularly since many students are unaware of the full scope of data collected, analyzed, and retained [23]. Swartz and McElroy find that heightened security concerns around these tools reduce student participation out of fear of being surveilled, and argue that the normalization of constant algorithmic monitoring during education conditions students to accept comparable surveillance as an ordinary feature of the workplaces they will later enter [23],

echoing Zuboff’s broader account of how routine data extraction becomes naturalized once it is embedded within an institution people cannot easily opt out of [6].

4.11 AI-Driven Anti-Money Laundering Systems

Financial institutions and regulators increasingly use AI to automatically detect suspicious transactions, identify complex money laundering schemes, and support compliance with anti-money laundering (AML) legislation [24]. The stakes of getting this wrong are illustrated by the case of HSBC Bank, which was fined approximately \$2 billion by a U.S. regulator after failing to prevent Mexican drug trafficking organizations from laundering funds through its banking channels, one of the clearest illustrations of why financial institutions have turned to AI-driven detection at scale [24].

4.11.1 Real-World Deployments & AI Techniques Employed

AML systems use machine learning techniques such as random forests, gradient boosting, support vector machines, and graph neural networks (GNNs) to analyze large volumes of transactional data, including payment amounts, the origin and destination of payments, timestamps, and relationships between accounts [24]. Financial institutions such as banks and other credit-granting organizations use these systems to detect risks, transactions, and possible money launderers, with GNNs in particular used to map relationships between accounts, entities, and transactions in order to surface hidden laundering networks [24]. Research interest in this application of AI has grown rapidly, with bibliometric analysis of the Scopus database showing a 14.33% annual growth rate in publications on AI and machine learning for combating illegal financial operations [25]. The growth of cryptocurrencies and decentralized finance presents a distinct technical challenge: blockchain-based transactions mask traditional laundering indicators while their pseudonymous, instantaneous, and cross-border character allows sophisticated schemes to evade conventional detection [24].

4.11.2 Accountability and Privacy Issues

Effective AI-driven AML requires financial institutions and regulators to access a wide range of financial data on individuals and businesses, giving these systems the practical capacity to conduct population-level surveillance of economic activity [24]. The use of graph-based neural networks to map relationships between accounts creates permanent records of financial interactions regardless of whether those interactions are ultimately found to be legitimate or suspicious, meaning that AI-driven AML provides government agencies and financial institutions with a capacity for financial surveillance at a scale and depth not previously achievable [24].

4.11.3 Democratic, Ethical and Governance Concerns

Historical data used to train AML systems reflects prior enforcement priorities and investigative decisions; where particular communities, nationalities, or business types have previously been subject to greater scrutiny, the resulting algorithms tend to reproduce that pattern by disproportionately flagging transactions involving those same groups [24]. Systematic mapping

of the AML literature identifies transparency and explainability as a persistent, unresolved gap: many of the models used in this domain operate as effective black boxes whose decision logic is not transparent even to their developers, which conflicts directly with the due process expectation that individuals and institutions be told why a specific trans- action was flagged [24]. Bibliometric analysis of the field further shows that ethical concerns, privacy, and bias remain comparatively underexplored relative to the technical detection literature, despite being identified as crucial to the fair and effective use of AI and ML in this domain [25]. This opacity produces a familiar accountability gap, echoing the same structural problem Busuioc identifies across algorithmic public-sector tools: current AML regulatory structures, including bodies such as the EU's Anti-Money Laundering Authority (AMLA), are still adapting to the pace of AI development, leaving uncertainty over who is accountable when an AI-driven system fails, whether by missing genuine laundering activity of the kind that went undetected at HSBC, or by wrongly freezing a legitimate account [24, 8]. Those most harmed by false positives tend to be the individuals with the least access to resources needed to appeal an algorithmic determination or to find an alternative financial institution once their account is closed [24].

Table 1: I Surveillance Solutions in Government Operations

Solution Name	Description	DOI/Link	Reference	Real-World Deployments & AI Techniques	Accountability & Privacy Issues	Democratic, Ethical & Governance Concerns
Real-time Facial Recognition (FRT) [4]	Real-time identity validation in public spaces for law enforcement and suspect identification.	10.1016/j.techsoc.2022.102137	[4]	Computer Vision, Biometric Matching, Deep Learning.	Loss of anonymity in public spaces, risk of ubiquitous mass dataveillance.	Potential for discrimination due to algorithmic bias (race/gender), “chilling effect” on democratic values.
Digital Contact Tracing Apps [4]	Applications for tracking contacts and limiting the spread of pandemics (e.g., COVID-19).	10.1016/j.techsoc.2022.102137	[4]	Risk Scoring Algorithms (calculating infection probability).	Centralized vs. Decentralized data storage, GPS vs. Bluetooth usage.	Voluntary vs. Mandatory adoption, exclusion of vulnerable groups without smartphones (e.g., elderly).
Automated Gunshot Detection [13]	Acoustic sensor networks that detect gunfire in urban areas and alert police in real time.	https://doi.org/10.36646/mjlr.55.4.bang	[13, 14]	Acoustic Sensing, Triangulation, Human-in-the-loop Review.	Incidental audio capture of speech, disproportionate sensor coverage of minority neighborhoods.	Low confirmation rate of alerts, reinforcement of racialized over-policing, unclear vendor accountability for false alerts.
Crowd Safety Manager [16]	Networked cameras and sensors that monitor crowd density and public activity in urban areas.	https://doi.org/10.1287/inte.2016.0860	[16, 17]	Computer Vision, Crowd Counting, Pattern Recognition Analytics.	Weak oversight relative to expanding scope, lack of archived outputs for independent audit.	Function creep from counterterrorism to general policing, minimal external accountability.
Autonomous Vehicle Sensor Fusion [18]	Fleet vehicle cameras used for navigation are also accessed by police as investigative footage.	10.3390/s23063335	[18, 19]	Deep Neural Networks (DNN), 4D Object Detection, Sensor Fusion.	Continuous recording of public space by a mobile fleet; inconsistent data retention practices.	Police access to footage without standardized legal process; oversight lagging practice.
Predictive Policing (Pred-Pol) [1]	Algorithmic prediction of crime locations or suspects using historic arrest data.	10.1111/j.1740-9713.2016.00960.x	[1, 20, 21]	Predictive Analytics, Pattern Recognition.	Aggregated risk scoring of individuals absent specific suspicion.	Algorithmic bias reproducing over-policing; runaway feedback loops; opaque accountability.
Monitoring a Smart City [2]	Smart City sensors (traffic, lights) repurposed for tracking citizen behavior.	10.1007/s43681-022-00196-y	[2, 1, 17]	Topic Modeling, IoT Sensor Fusion.	Function Creep: Benign data collection turns into invasive monitoring.	The “Ambiguity” Problem: Blurs the line between urban management and state control.
Udbetaling Danmark (UDK) [22]	Centralized authority merging data (tax, housing) to detect social benefit fraud.	10.1080/1369118X.2021.1934069	[22, 9]	Data Matching & Merging, Fraud Profiling Models.	Ubiquitous Dataveillance: Citizens’ lives transparent to state without suspicion (Section 12a).	Erosion of Trust: Vulnerable citizens are disproportionately surveilled compared to wealthy tax-evaders.
AI at Borders [3]	Integration of AI tools by EU border agencies (Frontex, EUROSUR, iBorderCtrl) to manage migration flows.	10.14746/rie.2025.19.8	[3, 8]	Predictive Analytics, Biometric Surveillance, Affect Recognition.	High risk to transparency due to “black box” algorithms; no framework governing cross-border data sharing.	Risk of discrimination against migrants; pseudoscientific lie detection; unclear liability for automated decisions.
The “Academicon” (Proctoring) [23]	Surveillance ecosystem in higher education (Proctorio, Social Sentinel).	(Journal)	[23]	Gaze Detection, Keystroke Dynamics, Geofencing.	Invasion of the Home (room scans); Data commodification by vendors.	Disability Bias (neurodivergent students flagged); Conditions students to accept constant monitoring.
AI-Driven AML Surveillance [24]	Monitoring banking transactions to detect money laundering, illustrated by the HSBC AML failure.	https://doi.org/10.1080/08839514.2024.2344415	[24, 25]	Graph Neural Networks, Anomaly Detection.	Financial panopticon: population-level surveillance via continuous monitoring.	Financial exclusion (false positives freezing assets); black box accountability (hard to appeal).

5

Discussion

5.1 Interpretation of Findings

A cross-domain systematic evaluation of 7 governmental domains and 11 AI surveillance tools identified a common problem: the development of the technology has occurred without an adequate legal framework to govern its use, thereby creating a gap in regulation and therefore a risk to the privacy and ethical rights of citizens. Each of the solutions was developed to serve a specific purpose (public health contact tracing, financial fraud detection etc.), however when viewed through a cross-cutting lens the similar patterns in which the technology is being used, justified and challenged were apparent [2].

A recurring theme is the phenomenon of function creep, whereby systems originally designed for narrow, benign purposes are incrementally expanded to perform broader surveillance functions. Smart city sensors deployed for traffic management are repurposed for behavioral monitoring; crowd safety systems are extended to track demographic groups; welfare fraud detection models incorporate social media data far beyond their original mandate [22]. This pattern suggests that the primary risk of AI surveillance lies not only in intentionally intrusive systems, but in the gradual, often opaque expansion of systems that citizens initially accepted for legitimate purposes.

Sensors are the common thread that incorporates an individual into every solution surveyed, whether located on the solution itself (e.g., a wearable device) or in the surrounding environment (e.g., a location-based smartphone application or a fixed street camera). Once a sensor detects that an individual has entered a specific area, or is engaging in a particular behaviour, the solution responds automatically. Figures 2 and 3 summarise each solution across

the three analytical axes examined in Chapter 4, and the patterns identified within each axis are discussed in turn below

5.1.1 Real-World Deployments & AI Techniques Employed

Analysis indicates that 9 of the 11 solutions utilize Computer Vision and/or Deep Learning techniques. Specifically, Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) allow for visual analysis to be performed automatically within the realm of facial recognition, gunshot detection and crowd monitoring [18]. Additionally, biometric matching technologies are utilized in four of the solutions (facial recognition, border control, and contact tracing). Machine learning algorithms are used to perform risk assessments and identify patterns in predictive policing and anti-money laundering systems. Furthermore, sensor fusion techniques that integrate LiDAR, radar, and camera feeds have emerged as an advanced technology in autonomous vehicle surveillance [18]. In every case, the technique is paired with a documented, real-world deployment, whether a specific city, police department, or agency, so that the analytical axis captures not only technical capability but evidence of that capability actually being put into practice.

5.1.2 Accountability and Privacy Issues

Three main privacy considerations appear throughout the solutions analysed. First, 7 of 11 applications cause a loss of anonymity in public areas, primarily those utilising facial recognition or continuous video surveillance [4]. Second, whether data is stored in a centralised or decentralised manner is a major design consideration: contact tracing apps illustrate both approaches, with GPS-based systems creating centralised databases of location history while Bluetooth proximity detection creates decentralised records of who has been near whom [4]. Third, many smart city platforms, autonomous vehicles, and educational proctoring systems continuously monitor user activity, creating permanent digital records without temporal limits. These privacy risks are compounded by an accountability deficit: law enforcement solutions in particular demonstrate persistent black-box accountability problems, where the processes underlying an algorithmic decision are opaque even to the officials relying on them, limiting citizens' ability to dispute automated determinations. Financial surveillance systems illustrate the same problem from another angle, with high false-positive rates capable of freezing individuals' assets with no clear route of appeal, underscoring the need for human oversight of high-stakes automated decisions.

5.1.3 Democratic, Ethical and Governance Concerns

Algorithmic bias is the most prominent concern within this axis, referenced in 8 of the 11 solutions. Documented racial and gender biases in facial recognition systems produce false positive rates disproportionate to minority populations [3]; predictive policing algorithms continue to support the long-standing practice of over-policing marginalised communities because they are trained on biased arrest data [1]; and educational proctoring systems flag neurodivergent students as suspicious due to atypical eye movements, a form of disability discrimination [23].

Alongside bias, the widespread normalisation of surveillance culture is evident across domains, with several solutions conditioning users to treat constant monitoring as a normal and required condition of participation, whether in education, employment, or public life. Public health applications illustrate a related democratic tension between voluntary and mandatory use: compulsory installation of AI solutions raises ethical questions about state coercion [4], questions that recur, in different form, wherever a surveillance capability expands without a corresponding expansion of democratic consent.

5.2 Common Democratic, Security and Ethical Considerations Across Solutions

Despite operating in distinct governmental domains, the eleven solutions examined in this review share a set of recurring democratic, security, and ethical concerns that transcend any single application context. Because each case study in Chapter 4 was analysed along the same three axes, Real-World Deployments & AI Techniques Employed, Accountability and Privacy Issues, and Democratic, Ethical and Governance Concerns, the patterns identified below are not new findings so much as the cross-case aggregation of what recurred, case after case, within the latter two axes specifically. Read together, they substantiate the governance gap framework set out in the Background: each theme below is a different empirical face of the same underlying gap between what these systems are technically capable of and what institutions have been built to oversee.

5.2.1 Algorithmic Bias and Discriminatory Outcomes

Algorithmic bias emerges as the most pervasive concern within the Democratic, Ethical and Governance Concerns axis, documented across the majority of solutions examined. Facial recognition systems produce false positive rates of up to 35% for dark-skinned women compared to 0.2% for white men [3]. Predictive policing systems trained on historically biased arrest data continue to over-police minority neighbourhoods, a pattern formally shown to produce runaway feedback loops [20, 21]. AI proctoring systems disproportionately flag neurodivergent students as suspicious [23]. Border biometric systems reject African and Middle Eastern travellers at rates exceeding 20% at some checkpoints [3]. Welfare fraud detection models disproportionately target low-income and immigrant communities [22], and gunshot detection alerts concentrate policing attention in the same historically over-policed neighbourhoods where sensors are installed [14]. Across all these cases, the common mechanism is the same: algorithms trained on historically unequal data reproduce and amplify existing social inequalities at scale and speed that human oversight cannot easily detect or correct, precisely the dynamic Lyon and Zuboff describe as the normalisation of data extraction within institutions citizens cannot easily opt out of [5, 6].

5.2.2 Opacity and Black-Box Accountability

A second systemic concern, falling within the Democratic, Ethical and Governance Concerns axis, is the opacity of algorithmic decision-making. In the majority of solutions examined, including predictive policing, AML surveillance, border risk assessment, and welfare fraud detection, the

criteria by which individuals are flagged, scored, or denied access to services are not disclosed to those affected [24, 1]. This opacity creates a fundamental accountability gap of exactly the kind Busuioc identifies as structural to algorithmic public-sector tools: because responsibility is distributed across vendors, operating agencies, and the officials who act on a system's output, no single actor can be clearly held to account when a determination proves wrong [8]. The black-box nature of these systems is particularly consequential in high-stakes domains such as border control and financial regulation, where automated determinations directly affect individuals' freedom of movement and access to economic resources.

5.2.3 Erosion of Anonymity and the Right to Private Life

Within the Accountability and Privacy Issues axis, the majority of solutions involve the continuous or large-scale collection of personal data in contexts where individuals have not provided meaningful consent. Real-time facial recognition, smart city monitoring, autonomous vehicle sensor fusion, and digital contact tracing apps all operate by capturing data about every individual within a surveilled environment, regardless of whether any specific suspicion exists [4, 18]. This structural feature, mass data collection as a precondition for system operation, fundamentally alters the relationship between citizen and state, transforming public space into a monitored environment and eroding the practical possibility of anonymous participation in civic life, the shift from bounded to diffuse surveillance that Lyon identifies as characteristic of contemporary surveillance assemblages [5].

5.2.4 Normalisation of Surveillance Culture

A less immediately visible but deeply significant concern, again within the Democratic, Ethical and Governance Concerns axis, is the progressive normalisation of surveillance as an ordinary feature of institutional life. AI proctoring systems condition students to accept constant monitoring as a prerequisite for educational participation [23]. Smart city platforms and urban sensor networks such as the NYPD's Domain Awareness System frame comprehensive behavioural monitoring as a public safety or service improvement, obscuring the extent to which their scope has expanded beyond their original, narrower mandate [17]. In each case, surveillance is embedded within legitimate institutional functions in ways that make resistance difficult and opt-out effectively impossible, the process Zuboff describes as the routinisation of data extraction once it is embedded within an institution people cannot easily avoid [6]. The cumulative effect across domains is the gradual establishment of a social norm in which being monitored is the default condition of participation in public, educational, financial, and professional life.

5.2.5 Democratic Accountability and Public Contestation

Beyond bias and opacity, several solutions raise a distinct concern about democratic legitimacy: whether citizens and their elected representatives retain meaningful control over surveillance infrastructure once it is deployed. Facial recognition has provoked direct democratic pushback, with cities including San Francisco, Oakland, and Somerville passing ordinances banning government use of the technology, and residents in other contexts taking evasive action, such as

Hong Kong protesters disabling facial recognition on their phones, when no legal channel for contestation was available [1]. Border AI systems and welfare fraud detection models are typically expanded through procurement decisions and administrative rulemaking rather than legislative debate, meaning the public has little opportunity to weigh in before the systems' scope is set [3, 22]. The absence of universally applicable human rights frameworks for many of these technologies allows both governments and private vendors to unilaterally expand their use, disproportionately increasing the vulnerability of minorities, children, and other marginalized groups who have the least power to contest that expansion [2]. This pattern is precisely what Danaher identifies as the legitimacy problem at the heart of algorithmic governance: when systems are deployed and iterated faster than democratic oversight mechanisms can respond, the resulting authority over people's lives is never meaningfully sanctioned by the public it affects [7].

5.2.6 Absence of Adequate Regulatory Frameworks

Finally, across the solutions examined, a consistent finding within the Democratic, Ethical and Governance Concerns axis is the absence of regulatory frameworks proportionate to the risks posed. Existing legal instruments, including the GDPR in the European context, were not designed with the scale, speed, and opacity of AI-driven surveillance in mind [3]. Border AI systems operate under procurement frameworks that bypass democratic scrutiny. Welfare surveillance systems are expanded through administrative decisions rather than legislative processes. Financial AML systems are governed by sectoral regulations that do not address algorithmic accountability, and the EU's own Anti-Money Laundering Authority is still adapting to the pace of the technology it is meant to oversee [24]. The result is a landscape in which the deployment of AI surveillance consistently precedes the development of the governance mechanisms needed to ensure that such deployment respects fundamental rights, the governance gap named at the outset of this chapter and traceable, case by case, to the specific legitimacy and accountability deficits that Danaher and Busuioc theorise [7, 8, 2, 1]

5.3 Emerging Risks in AI Surveillance

The eleven cases analysed in this chapter document surveillance capabilities that are already deployed. A governance framework built only around present capabilities, however, risks being obsolete on arrival, since the technologies underpinning these systems are themselves changing quickly. Four categories of emerging risk deserve particular attention because each widens the governance gap rather than closing it.

5.3.1 Generative AI, Synthetic Identities, and Deepfake

Generative AI substantially lowers the cost, time, and skill required to fabricate the identity documents and biometric material on which several of the systems examined here depend. The U.S. Financial Crimes Enforcement Network has formally alerted financial institutions that criminals are using generative AI to create falsified documents, photographs, and videos specifically in order to circumvent customer identification and verification processes, and has advised institutions that deepfake-produced identity documents may now pass checks that were

designed on the assumption that document forgery required specialist skill [26]. The implication cuts across multiple cases in this chapter. Border biometric systems, AML customer verification, and facial recognition databases all treat a captured biometric or document as evidence of identity; if that evidence can be synthesised at scale, the evidentiary foundation of these systems weakens precisely as institutions become more dependent on them. This creates a distinctive governance problem: the same institutions are simultaneously deploying AI surveillance to detect wrongdoing and becoming more vulnerable to AI-generated deception, with no clear allocation of responsibility for the resulting errors.

5.3.2 Adversarial Attacks and Cybersecurity Vulnerabilities

Deployed surveillance systems are not merely imperfect; they are actively attackable. A comprehensive survey of adversarial attacks against facial recognition documents that deep learning based systems exhibit a systematic vulnerability to input images carrying imperceptible, or perceptible but natural-looking, perturbations that drive the model to incorrect predictions, with demonstrated attacks including printed stickers, patterned eyewear, and projected light [27]. Because these attacks target the model rather than the surrounding infrastructure, conventional cybersecurity controls do not detect them, and the opacity documented throughout this chapter compounds the problem: an agency that cannot explain why its system produced a given output is also poorly positioned to determine whether that output was manipulated. The governance consequence is that accuracy claims made by vendors, already difficult to verify independently in the gunshot detection and facial recognition cases, become still less reliable once adversarial manipulation is treated as a realistic operating condition rather than a laboratory curiosity.

5.3.3 Interconnected Surveillance Ecosystems and Cross-Border Data Sharing

Several of the cases examined here are no longer standalone systems. The NYPD's Domain Awareness System integrates CCTV, licence plate readers, chemical sensors, and gunshot detection feeds into a single analytical platform [16], autonomous vehicle fleets have become an ad hoc source of investigative footage for police departments [19], and EU border databases covering visa applications, asylum claims, passenger records, and criminal records are increasingly interconnected through algorithms with little transparency and no comprehensive legal framework governing data sharing among member states [3]. Interconnection multiplies the governance gap rather than adding to it linearly: each system was individually authorised for a bounded purpose, but the combined platform performs a form of population monitoring that no single authorisation ever contemplated, and no single oversight body has jurisdiction over the whole.

5.3.4 Autonomous Operation and the Limits of Human Oversight

Finally, the speed at which these systems operate increasingly strains the human oversight that governance frameworks rely upon as a safeguard. Gunshot detection dispatches alerts within thirty to forty-five seconds [13], predictive policing models update continuously as new arrest data arrives [20], and welfare profiling systems scan entire benefit populations on a recurring schedule

[22]. Where a human reviewer is nominally in the loop, the volume and pace of automated output can reduce that role to procedural confirmation rather than substantive review. This matters for governance design: requirements for human oversight are meaningful only if the reviewing officer has both the time and the information needed to disagree with the system, and neither condition can be assumed simply because a human is formally positioned in the decision chain.

Taken together, these developments suggest that governance frameworks for AI surveillance should be principles-based rather than technology-specific, since rules written around the capabilities of currently deployed systems will not extend to the systems that succeed them. They also strengthen the case for the horizon-scanning and periodic reauthorisation requirements set out in the recommendations that follow.

6

Conclusion

The systematic literature review reveals important contradictions between the efficiency of operation and the protection of individual rights of AI-driven government surveillance. Computer vision and deep learning are included in nearly all 11 of the technical implementations of AI-based government surveillance systems studied in the review; yet algorithmic bias and loss of privacy were identified as common issues in many of the systems studied. Also, the review's results reveal that at present most of the current deployments of AI-based surveillance systems lack the appropriate governance frameworks, transparency and democratic oversight [3].

Based on the findings of this systematic review of the literature, six governance recommendations follow, each linked to a specific deficit identified in the preceding analysis.

First, algorithmic impact assessments should be a mandatory precondition for procurement rather than a post-deployment formality. The recurring pattern across the smart city, border, and welfare cases is that surveillance capability is acquired through procurement decisions taken before any rights assessment occurs, which forecloses meaningful scrutiny at the only point where a deployment can still be reconsidered [3, 22].

Second, independent algorithmic auditing should be institutionalised in bodies possessing both technical expertise and legal authority to compel disclosure. The gunshot detection and predictive policing cases show that vendor-generated accuracy claims are difficult to verify independently, and that where independent audits have occurred, they have produced findings substantially at odds with vendor representations [13, 20].

Third, accessible redress mechanisms must allow citizens to contest automated determinations and obtain a substantive explanation of the basis on which they were flagged. The

welfare surveillance and AML cases demonstrate that those most harmed by false positives are typically those least equipped to appeal them [22, 24].

Fourth, new surveillance deployments should require explicit legislative authorisation rather than proceeding through administrative rulemaking or procurement. The Domain Awareness System's migration from counterterrorism tool to general policing platform under an unchanged privacy policy illustrates how function creep proceeds precisely where no legislative reauthorisation is required [17].

Fifth, transparency obligations should extend beyond acknowledging that a system exists to publishing its operational parameters, error rates, and demographic performance differentials, together with the retention and disclosure practices governing the data it collects. The autonomous vehicle case, where retention periods and law enforcement disclosure practices vary between operators with no consistent standard, shows how the absence of such requirements leaves both citizens and oversight bodies unable to assess what is actually happening [19].

Sixth, participatory governance structures incorporating civil society, data protection authorities, and affected communities should be established as permanent institutional features rather than consultative exercises convened after deployment, replacing the technocratic model that currently governs most AI surveillance systems. Sustaining these six measures over time additionally requires investment in institutional capacity: oversight bodies, courts, and procurement officials cannot meaningfully scrutinise systems whose technical operation they lack the expertise to interrogate, and the emerging risks discussed in Section 5.3 suggest that this capacity must be accompanied by periodic reauthorisation and horizon-scanning obligations if governance is to keep pace with the technology it is meant to constrain. Future research is needed to study the long-term effects of normalized surveillance on democratic participation, particularly in marginalized communities where surveillance is experienced at a greater level due to demographic characteristics. Comparative studies of regulatory frameworks employed in various jurisdictions could also contribute to determining effective governance models that both promote national security objectives and protect fundamental rights. Lastly, future research will necessitate interdisciplinary collaborations between researchers in fields such as computer science, law, sociology, etc., to develop ethical technology systems that consider ethics during design phases and do not treat ethics as post-deployment constraints

References

- [1] S. Feldstein, The global expansion of AI surveillance. Carnegie Endowment for International Peace Washington, DC, 2019, vol. 17, no. 9.
- [2] T. Saheb, “Ethically contentious aspects of artificial intelligence surveillance: a social science perspective,” *AI and Ethics*, vol. 3, no. 2, pp. 369–379, 2023.
- [3] C. Imbrosio Filho, “Artificial intelligence in european border policing: Legal challenges, migration governance, and security sector reform,” *Rocznik Integracji Europejskiej*, no. 19, pp. 133–156, 2025.
- [4] C. Fontes, E. Hohma, C. C. Corrigan, and C. L’utge, “Ai-powered public surveillance systems: why we (might) need them and how we want them,” *Technology in Society*, vol. 71, p. 102137, 2022.
- [5] D. Lyon, “Surveillance studies: An overview,” 2007.
- [6] S. Zuboff, “The age of surveillance capitalism,” in *Social theory re-wired*. Routledge, 2023, pp. 203–213.
- [7] J. Danaher, M. J. Hogan, C. Noone, R. Kennedy, A. Behan, A. De Paor, H. Felzmann, M. Haklay, S.-M. Khoo, J. Morison et al., “Algorithmic governance: Developing a research agenda through the power of collective intelligence,” *Big data & society*, vol. 4, no. 2, p. 2053951717726554, 2017.
- [8] M. Busuioc, “Accountable artificial intelligence: Holding algorithms to account,” *Public administration review*, vol. 81, no. 5, pp. 825–836, 2021.
- [9] M. H. Moore, *Creating public value: Strategic management in government*. Harvard university press, 1997.
- [10] E. L. Trist et al., *The evolution of socio-technical systems*. Ontario Quality of Working Life Centre Toronto, 1981, vol. 2.
- [11] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, “Lessons from applying the systematic literature review process within the software engineering domain,” *Journal of system and software*, vol. 80, no. 4, pp. 571–583, 2007.
- [12] R. C. Nickerson, U. Varshney, and J. Muntermann, “A method for taxonomy development and its application in information systems,” *European journal of information systems*, vol. 22, no. 3, pp. 336–359, 2013.
- [13] H. Gee, “‘bang!’: Shotspotter gunshot detection technology, predictive policing, and measuring terry’s reach,” *University of Michigan Journal of Law Reform*, vol. 55, no. 4, pp. 767–812, 2022.

- [14] E. L. Piza et al., “Space-time association between gunshot detection alerts, calls for service, and police enforcement in Chicago: Differences across citizen race and incident type,” *Journal of Quantitative Criminology*, vol. 41, no. 1, pp. 53–74, 2025.
- [15] M. L. Doucette et al., “Impact of shotspotter technology on firearm homicides and arrests among large metropolitan counties: A longitudinal analysis, 1999–2016,” *Journal of Urban Health*, vol. 98, no. 5, pp. 609–621, 2021.
- [16] E. S. Levine, J. Tisch, A. Tasso, and M. Joy, “The New York City Police Department’s domain awareness system,” *Interfaces*, vol. 47, no. 1, pp. 70–84, 2017.
- [17] A. Diaz, “New York City Police Department surveillance technology,” Brennan Center for Justice Tech. Rep., 2019.
- [18] M. Hasanujjaman, M. Z. Chowdhury, and Y. M. Jang, “Sensor fusion in autonomous vehicle with traffic surveillance camera system: detection, localization, and AI networking,” *Sensors*, vol. 23, no. 6, p. 3335, 2023.
- [19] TechCrunch, “Police want robotaxi video footage to help solve crimes,” TechCrunch, 2023, reporting on Bloomberg’s 2023 investigation into police requests for Waymo and Cruise footage. [Online]. Available: <https://techcrunch.com/2023/06/30/police-want-robotaxi-video-footage-to-help-solve-crimes>
- [20] K. Lum and W. Isaac, “To predict and serve?” *Significance*, vol. 13, no. 5, pp. 14–19, 2016.
- [21] D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian, “Runaway feedback loops in predictive policing,” in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, ser. PMLR, vol. 81, 2018, pp. 160–171.
- [22] R. F. Jørgensen, “Data and rights in the digital welfare state: the case of Denmark,” *Information, Communication & Society*, vol. 26, no. 1, pp. 123–138, 2023.
- [23] M. Swartz and K. McElroy, “The “academicon”: AI and surveillance in higher education,” *Surveillance & Society*, vol. 21, no. 3, pp. 276–281, 2023.
- [24] H. U. Khan, M. Z. Malik, and S. Nazir, “Identifying the AI-based solutions proposed for restricting money laundering in financial sectors: Systematic mapping,” *Applied Artificial Intelligence*, vol. 38, no. 1, p. 2344415, 2024.
- [25] S. Lyeonov, V. Draskovic, Z. Kubaščíková, and V. Fenyves, “Artificial intelligence and machine learning in combating illegal financial operations: Bibliometric analysis,” *Human Technology*, vol. 20, no. 2, pp. 325–360, 2024.
- [26] Financial Crimes Enforcement Network, “Fincen alert on fraud schemes involving deepfake media targeting financial institutions,” U.S. Department of the Treasury, Tech. Rep. FIN-2024-Alert004, 2024. [Online]. Available: <https://www.fincen.gov/system/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf>
- [27] F. Vakhshiteh, A. Nickabadi, and R. Ramachandra, “Adversarial attacks against face recognition: A comprehensive study,” *IEEE Access*, vol. 9, pp. 92 735–92 756, 2021.

7

Appendix.

Table A1 presents the description, source, and AI techniques employed for each of the eleven surveillance solutions analysed in this chapter; Table A2 presents the associated accountability, privacy, and democratic, ethical and governance concerns. The Solution Name column is common to both tables, allowing entries to be cross-referenced between them. Together, these tables constitute the coded dataset produced through the content analysis described in Section 3.4.

Table 2: AI Surveillance Solutions — Sources and Technical Characteristics

Solution Name	Description	DOI/Link	Source(s)	AI Techniques Employed
Real-Time Facial Recognition (FRT)	Real-time identity validation in public spaces for law enforcement and suspect identification .	10.1016/j.techsoc.2022.102137	Fontes et al. 2022; Feldstein 2019	Computer Vision, Biometric Matching, Deep Learning.
Digital	Applications	10.1016/j.techsoc.2022.102137	Fontes et al.	Risk Scoring

Διπλωματική εργασία: AI Surveillance in Government: Governance, Accountability, and Democratic Challenges in the Digital State

Contact Tracing Apps	for tracking contacts and limiting the spread of pandemics (e.g., COVID-19).		2022	Algorithms (calculating infection probability).
Automated Gunshot Detection	Acoustic sensor networks that detect gunfire in urban areas and alert police in real time.	10.36646/mjlr.55.4.bang	Gee 2022; Piza et al. 2025	Acoustic Sensing, Triangulation , Human-in-the-loop Review.
Crowd Safety Manager	Networked cameras and sensors that monitor crowd density and public activity in urban areas.	10.1287/inte.2016.0860	Levine et al. 2017; Diaz 2019	Computer Vision, Crowd Counting, Pattern Recognition Analytics.
Autonomous Vehicle Sensor Fusion	Fleet vehicle cameras used for navigation are also accessed by police as investigative footage.	10.3390/s23063335	Hasanujjaman et al. 2023; TechCrunch 2023	Deep Neural Networks (DNN), 4D Object Detection, Sensor Fusion.
Predictive Policing (PredPol)	Algorithmic prediction of crime locations or suspects using historic arrest data.	10.1111/j.1740-9713.2016.00960.x	Feldstein 2019; Lum & Isaac 2016; Ensign et al. 2018	Predictive Analytics, Pattern Recognition.

Διπλωματική εργασία: AI Surveillance in Government: Governance, Accountability, and Democratic Challenges in the Digital State

Monitoring a Smart City	Smart City sensors (traffic, lights) repurposed for tracking citizen behavior.	10.1007/s43681-022-00196-y	Saheb 2023; Feldstein 2019; Diaz 2019	Topic Modeling, IoT Sensor Fusion.
Udbetaling Danmark (UDK)	Centralized authority merging data (tax, housing) to detect social benefit fraud.	10.1080/1369118X.2021.1934069	Jørgensen 2023; Moore 1997	Data Matching & Merging, Fraud Profiling Models.
AI at Borders	Integration of AI tools by EU border agencies (Frontex, EUROSUR, iBorderCtrl) to manage migration flows.	10.14746/rie.2025.19.8	Imbrosio Filho 2025; Busuioc 2021	Predictive Analytics, Biometric Surveillance, Affect Recognition.
The "Academicon" (Proctoring)	Surveillance ecosystem in higher education (Proctorio, Social Sentinel).	10.24908/ss.v21i3.16105	Swartz & McElroy 2023	Gaze Detection, Keystroke Dynamics, Geofencing.
AI-Driven AML Surveillance	Monitoring banking transactions to detect money laundering, illustrated by the HSBC AML failure.	10.1080/08839514.2024.2344415	Khan et al. 2024; Lyeonov et al. 2024	Graph Neural Networks, Anomaly Detection.

Solution Name	Accountability & Privacy Issues	Democratic, Ethical & Governance Concerns
Real-Time Facial Recognition (FRT)	Loss of anonymity in public spaces, risk of ubiquitous mass dataveillance.	Potential for discrimination due to algorithmic bias (race/gender), "chilling effect" on democratic values.
Digital Contact Tracing Apps	Centralized vs. decentralized data storage, GPS vs. Bluetooth usage.	Voluntary vs. mandatory adoption; exclusion of vulnerable groups without smartphones (e.g., elderly).
Automated Gunshot Detection	Incidental audio capture of speech, disproportionate sensor coverage of minority neighborhoods.	Low confirmation rate of alerts, reinforcement of racialized over-policing, unclear vendor accountability for false alerts.
Crowd Safety Manager	Weak oversight relative to expanding scope, lack of archived outputs for independent audit.	Function creep from counterterrorism to general policing, minimal external accountability.
Autonomous Vehicle Sensor Fusion	Continuous recording of public space by a mobile fleet; inconsistent data retention practices.	Police access to footage without standardized legal process; oversight lagging practice.
Predictive Policing (PredPol)	Aggregated risk scoring of individuals absent specific suspicion.	Algorithmic bias reproducing over-policing; runaway feedback loops; opaque accountability.
Monitoring a Smart City	Function creep: benign data collection turns into invasive monitoring.	The "ambiguity" problem: blurs the line between urban management and state control.
Udbetaling Danmark (UDK)	Ubiquitous dataveillance: citizens' lives transparent to the state without suspicion.	Erosion of trust: vulnerable citizens are disproportionately surveilled compared to wealthier tax evaders.
AI at Borders	High risk to transparency due to "black box" algorithms; no framework governing cross-	Risk of discrimination against migrants; pseudoscientific lie detection; unclear liability for

Διπλωματική εργασία: AI Surveillance in Government: Governance, Accountability, and Democratic Challenges in the Digital State

	border data sharing.	automated decisions.
The "Academicon" (Proctoring)	Invasion of the home (room scans); data commodification by vendors.	Disability bias (neurodivergent students flagged); conditions students to accept constant monitoring.
AI-Driven AML Surveillance	Financial panopticon: population-level surveillance via continuous monitoring.	Financial exclusion (false positives freezing assets); black-box accountability, hard to appeal.