REVIEW OF AI RISKS IN THE EUROPEAN UNION PUBLIC SECTOR: TAXONOMY, GOVERNANCE ARCHITECTURE AND CASE STUDIES (2018–2025)

Ilinka Delipetreva
Faculty of Economics and Business Administration, University "St. Kliment Ohridski", Sofia, Bulgaria,
delipetreva.ilinka@gmail.com

Abstract: European public administrations increasingly deploy artificial intelligence (AI) services for social-benefit and service eligibility screening, fraud and anomaly detection, citizen-service triage, and multilingual document processing. Because these deployments operate inside a complex constitutional and administrative order, the principal risks are not purely technical: they also implicate legality and mandate, proportionality, due process and reasons-giving, equality and non-discrimination, privacy and data protection, cybersecurity, and democratic accountability. This paper reviews AI risks in the European Union (EU) public sector over 2018–2025 by analyzing technical, legal, and policy literature that connects observed risks in administrative practice to the evolving governance architecture. Methodologically, it triangulates EU-level evidence on adoption and capacity constraints, the EU "governance stack" centered on the Artificial Intelligence Act (AI Act) and complemented by the General Data Protection Regulation (GDPR), the Data Governance Act, the Data Act, the Digital Services Act (DSA), the Digital Markets Act (DMA), and Network and Information Security Directive 2 (NIS2), and four selected cases: the Netherlands' SyRI welfare-fraud analytics, the EU-funded iBorderCtrl border pilot, justice/probation automation debates with Estonia as an illustrative boundary-setting instance, and Madrid's use and experimentation with facial recognition in mobility/security contexts. The analysis produces a risk taxonomy spanning technical, organizational, legal-accountability, and sovereignty dimensions and maps it to lifecycle duties under the AI Act. The paper concludes that the EU governance architecture is conceptually coherent and comparatively advanced, but its effectiveness will depend on procurement, integrated impact assessments, and sustained capacity-building so that AI adoption delivers public value without shifting risk onto citizens or undermining trust. It also identifies recurring administrative pressure points—such as auditability, logging, human oversight, and vendor dependency—that shape whether lifecycle obligations translate into day-to-day safeguards.
Keywords: public-sector AI, EU AI Act, risk taxonomy, accountability.

1. INTRODUCTION

AI has moved from experimentation to operational use across European public administration. Authorities deploy AI and language technologies to automate document-heavy processes, detect anomalies in taxation or customs, support case triage, and manage multilingual workflows. EU mapping work indicates that public-sector AI is becoming infrastructure rather than a niche set of pilots, while adoption remains uneven across Member States, levels of government, and organizational capability (Tangi et al., 2022).

In the EU, these deployments collide with a distinctive legal environment. Public authorities are bound by legality, equal treatment, proportionality, procedural fairness, and the duty to provide reasons and enable review. AI therefore creates compound risk: technical performance interacts with law and organizational practice. A system may appear accurate yet be unlawful if it lacks a proper mandate, is disproportionate relative to available alternatives, cannot provide reasons that enable contestation, produces discriminatory effects through proxy variables, or depends on vendor controls that block meaningful audit (European Union, 2024).

EU regulation has responded by building a layered governance stack. The AI Act establishes a horizontal, risk-based regime with lifecycle duties and extensive obligations for high-risk systems relevant to many public-sector domains. GDPR continues to govern personal-data processing, including minimization, purpose limitation, and safeguards (European Union, 2016). The Data Governance Act and Data Act shape access, reuse, interoperability and switching, which affect both data quality and dependency (European Union, 2022a; European Union, 2023). DSA/DMA constrain certain platform risks and gatekeeper dependencies relevant to digital service delivery (European Union, 2022b). NIS2 extends cybersecurity governance and incident reporting obligations across many public entities and their ICT supply chains (European Union, 2022c).

The literature remains fragmented. Technical work often emphasizes bias, drift, and adversarial vulnerability. Legal work centers on lawful bases, proportionality, transparency, and contestability. Policy works often highlights procurement and capacity constraints. The paper analyzes an EU-specific synthesis that connects risks observed in administrative practice to the governance architecture and draws lessons from use cases for the 2018–2025 period (Tangi et al., 2022).

2. MATERIALS AND METHODS

This paper is a structured survey and synthesis. It proceeds by triangulation of three evidence streams and treats governance as a lifecycle problem: what must be justified and evidenced before deployment, controlled during operation, and monitored after deployment.

First, it uses EU-level evidence on public-sector AI adoption and capability constraints to situate risks in common contexts of use and recurring organizational bottlenecks. The analytic goal is not completeness but realism: which domains and organizational settings repeatedly surface the same risk dynamics, and which capacity gaps systematically undermine assurance.

Second, it analyses the EU governance stack as a set of constraints and duties that operationalize administrative-law principles for AI. The AI Act supplies the horizontal, risk-based architecture and lifecycle obligations for high-risk systems, including risk management, data governance, documentation, logging, human oversight, robustness, and post-market monitoring. GDPR provides the baseline for lawful processing, transparency, minimization, purpose limitation, and high-risk processing assessment duties. The Data Governance Act and Data Act address data access, reuse, and switching conditions relevant to AI supply chains and lock-in. DSA/DMA shape transparency and power asymmetries in platform-mediated environments that may host or intermediate public service delivery. NIS2 extends cybersecurity risk management and incident reporting expectations across many public entities and their ICT providers, increasingly relevant as AI becomes embedded in critical workflows.

Third, it analyzes four selected cases from public sources. SyRI produced a landmark judicial outcome on proportionality and safeguards in welfare analytics (District Court of The Hague, 2020). iBorderCtrl illustrates the limits of high-stakes inference claims and generated parliamentary scrutiny (European Parliament, 2020). The Estonia use case serve as boundary-setting instance in justice automation narratives, illustrating how governance may explicitly reject "replacement" framings for adjudicative functions (Fabri, 2024). Madrid's facial recognition experience is as an example of biometrics in everyday public environments, raising proportionality, transparency, and public acceptance issues (AlgorithmWatch, 2020).

Public sources frequently do not specify model architecture, training data composition, or workflow integration details. Rather than infer parameters, this paper treats missing technical detail as part of the governance problem because contestability and oversight rely on evidence that is often absent in the public record.

3. RESULTS

Risks in EU public-sector AI emerge from interactions among data, models, infrastructure, administrative routines, procurement chains, and legal constraints. A practical taxonomy for EU administration can be organized into four dimensions: technical, organizational, legal-accountability, and sovereignty/dependency. These dimensions map to the lifecycle logic of the AI Act and its surrounding instruments.

Technical risk includes bias and uneven error, opacity and limited interpretability, robustness failures, and cybersecurity threats. Bias in public administration often originates in "policy-shaped data," where labels and features reflect enforcement patterns and historical administrative practices, meaning errors can translate into unequal burdens. Opacity matters because administrative legitimacy depends on intelligible reasons, not merely outputs. Robustness and security matter because public-sector systems operate in adversarial environments, particularly in enforcement, border, and benefit contexts, and because language model deployments create new vectors for prompt injection and unintended disclosure. The AI Act addresses these concerns via high-risk duties on data governance, documentation, logging, accuracy/robustness, and cybersecurity, and by embedding assurance as a lifecycle expectation rather than a one-off check.

Organizational risk includes over-automation, deskilling, digital exclusion, and "paper compliance" without operational control. Decision support can become decision substitution through interface design and workload pressure, eroding professional judgment and creating institutions that cannot detect model failure. Digital exclusion can occur when automated front doors displace human channels, disadvantaging people with low digital literacy, disabilities, or language barriers. Paper compliance occurs when obligations are documented but not operationalized, often due to insufficient expertise, time, or procurement leverage. These organizational risks are directly relevant to human oversight duties under the AI Act, which require meaningful oversight in context rather than nominal sign-off. The fact that capability is uneven across administrations is documented in EU mapping and directly shapes practical compliance outcomes.

Legal and accountability risk includes mandate, purpose limitation, proportionality, transparency, redress, and responsibility across supply chains. GDPR anchors lawful processing, minimization, and purpose limitation where personal data are involved and underpins contestability. The AI Act adds lifecycle documentation and monitoring that can strengthen proportionality review by producing evidence about design and operation, but legality cannot be retrofitted if a system is scoped too broadly or designed to avoid scrutiny. Accountability becomes complex when

vendors, integrators, and cloud services control model changes, thresholds, or logs, while citizens experience a single decision-maker: the state. The governance stack responds by allocating obligations across providers and deployers, but effectiveness depends on procurement that secures real audit access and remediation rights.

Sovereignty and dependency risk arises from reliance on proprietary models, compute, and vendor-controlled interfaces that constrain auditability, continuity, and lawful control. The Data Act's switching-related provisions are relevant in practice where cloud dependency and vendor lock-in limit an administration's ability to enforce requirements, move workloads, or retain evidence such as logs and documentation across suppliers. DSA/DMA can matter where platform dominance shapes public-service delivery ecosystems and bargaining power in ways that affect transparency and continuity. For public administration, sovereignty is not an option, it is the capability to maintain lawful, auditable control over essential public functions and to exercise exit options when suppliers cannot meet obligations.

Together, these risks map onto a governance architecture intended to prevent "black box administration" by requiring justification, traceability, and accountability over time. Yet this architecture assumes deployer capacity. Where capacity is weak, risk migrates from systems to organizations, and citizens' rights depend on administrative happenstance rather than consistent standards.

4. CASE STUDIES

SyRI in the Netherlands illustrates welfare analytics colliding with proportionality and transparency requirements. SyRI linked administrative datasets and generated risk indicators intended to support fraud detection and enforcement targeting. A core governance tension was secrecy about risk indicators to protect the system's integrity, which limited transparency and contestability for affected citizens. In February 2020, the District Court of The Hague held that the legal framework underpinning SyRI violated Article 8 of the European Convention on Human Rights, reasoning that the interference with private life lacked sufficient safeguards relative to the breadth and opacity of processing (District Court of The Hague, 2020). SyRI is important because it demonstrates that legality and legitimacy depend on more than predictive performance: broad data linkage and opaque scoring can fail proportionality review even when the policy objective is legitimate.

iBorderCtrl illustrates the problem of evidentiary validity in high-stakes inference. iBorderCtrl was an EU-funded Horizon 2020 project developing an "intelligent portable border control system," including tools intended to support border processes and reduce delays. The project attracted scrutiny about reliability, false positives, and fundamental rights implications, reflected in a European Parliament question to the European Commission. The governance lesson is that certain tasks are intrinsically legitimacy sensitive. If the underlying construct claimed to be measured is unstable or context-dependent, then documentation and oversight cannot reliably convert the system into a lawful and legitimate administrative practice. In such cases, responsible governance can mean discontinuation rather than controlled deployment.

Justice and probation automation debates often concentrate on whether predictive scoring can or should influence decisions affecting liberty. Estonia is invoked in public discourse as a "robot judge" example, yet careful discussion in the European court administration literature emphasizes that such narratives were misleading and that Estonia did not introduce an AI system to replace judicial decision-making (Fabri, 2024). This boundary-setting is important because it shows that explicit limits and non-deployment can be mature governance choices where fairness, interpretability, and reasons-giving cannot be satisfied by proposed technologies.

Madrid's facial recognition experience illustrates proportionality and public acceptance challenges of biometrics in everyday public environments. Reporting described that Madrid South bus terminal deployed live facial recognition matching visitors against a suspect database and sharing information with police, while noting limited transparency and uncertainty about effectiveness. Regardless of the precise technical configuration, the governance lesson is stable: biometric deployments in public spaces create severe human rights impacts, including for individuals who are not meaningfully able to opt out, and they invite function creep once infrastructure exists.

5. DISCUSSION

Across the cases, four determinants recur. Proportionality and purpose limitation are practical tests of legality. SyRI collapsed because broad linkage combined with opacity and insufficient safeguards could not be justified (District Court of The Hague, 2020). Madrid illustrates that biometric systems in everyday contexts rapidly trigger proportionality pressure because less intrusive alternatives are visible and because the rights impact extends beyond consenting users. These cases align with the governance stack's emphasis on lifecycle evidence: the legal acceptability of a system depends on narrow scoping, contestable reasons, and operational safeguards rather than formal authorization alone (European Union, 2024).

Evidentiary validity is decisive in high-stakes inference. iBorderCtrl illustrates that weakly validated measurement claims generate legitimacy crises that compliance paperwork cannot solve. This points to a practical governance gate: before any high-stakes deployment, administrations should require a defensible theory of measurement and context-appropriate validation evidence and should treat "uncertain validity" as a reason to stop, not to proceed cautiously.

Explainability is condition of administrative legitimacy. The governance stack requires transparency, logging, and human oversight, but these must be translated into workflow realities in which staff can challenge outputs and preserve human responsibility. Estonia's boundary-setting approach reinforces a core principle: for significant-effect decisions, administrations should preserve human reasons-giving and avoid designs where AI becomes a de facto decision-maker, even if formally described as "advisory" (Fabri, 2024).

Capacity and procurement are binding constraints. The AI Act and related instruments presume deployers can evaluate documentation, maintain logs, monitor change, and manage incidents, yet adoption evidence indicates uneven maturity and capability across administrations. This implies that procurement and shared capacity models are not optional add-ons, they are prerequisites for uniform rights protection under harmonized rules. In practical terms, the most defensible public-sector AI deployments will be those supported by an integrated assurance package: a documented proportionality case, coordinated impact assessment and risk files, contracts that guarantee access to evidence and logs, meaningful operational oversight, and continuous monitoring and incident response.

6. CONCLUSIONS

AI is becoming an important component of EU public administration, but the risks it introduces are multi-dimensional: technical error and security failures interact with legality, proportionality, reasons-giving, transparency and accountability. The EU governance stack, centered on the AI Act and complemented by GDPR, data governance law, platform regulation, and NIS2, provides a coherent lifecycle architecture intended to prevent black-box administration. The case studies show a consistent pattern: where proportionality, transparency, and evidentiary validity are weak, legal and social legitimacy collapse. Where institutions define boundaries and preserve reasons-giving, non-deployment can itself be a mature governance choice. The decisive variable for the next phase is capacity. Europe can scale public-sector AI responsibly only by scaling assurance capacity alongside adoption, so that AI delivers public value without shifting risk onto citizens or eroding trust.

DISCLAIMER

The views expressed in this article are purely those of the authors and may not, under any circumstances, be regarded as an official position of the European Commission.

REFERENCES

AlgorithmWatch. (2020). Spain's largest bus terminal deployed live face recognition. https://algorithmwatch.org/en/spain-mendez-alvaro-face-recognition/

District Court of The Hague. (2020). Judgment on SyRI.

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

European Parliament & Council of the European Union. (2022a). Regulation (EU) 2022/868 of 30 May 2022 on European data governance (Data Governance Act).

European Parliament & Council of the European Union. (2022b). Regulation (EU) 2022/2065 of 19 October 2022 on a Single Market for Digital Services (Digital Services Act) and Regulation (EU) 2022/1925 of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act).

European Parliament & Council of the European Union. (2022c). Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive).

European Parliament & Council of the European Union. (2023). Regulation (EU) 2023/2854 of 13 December 2023 on harmonized rules on fair access to and use of data (Data Act).

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act).

Fabri, M. (2024). From court automation to e-justice and beyond in Europe. The International Journal for Court Administration, 15(2).

Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., & Pignatelli, F. (2022). AI Watch: European landscape on the use of artificial intelligence by the public sector (EUR 31088 EN). Publications Office of the European Union.