ANALYTICAL REPORT 05

FROM COMPLIANCE TO TRUST: OPERATIONALISING RIGHTS-BASED GOVERNANCE FOR AI-ENABLED INTELLIGENCE SUPPORT TECHNOLOGIES IN THE EU FCT DOMAIN

An External Expert Report prepared for ENACT

Main Authors
Antonio Landi

August 2026

ENACT.
European Network Against Crime and Terrorism

ABOUT ENACT

ENACT is a knowledge network focused on the fight against crime and terrorism (FCT). The network is funded under the Horizon Europe Framework Programme in Cluster 3 – Civil Security for Society. The project addresses the call topic HORIZON-CL3-2022-SSRI-01-02 'Knowledge Networks for Security Research & Innovation', aiming to collect, aggregate, process, disseminate and make the most of the existing knowledge in the FCT area.

The project aims to satisfy two major ambitions,

- Provide evidence-based support to the decision-makers in the EU research and innovation (R&I) ecosystem in the FCT domain, targeted explicitly at enabling more effective and efficient programming of EU-funded R&I for the fight against crime and terrorism.

- Act as a catalyst for the uptake of innovation by enhancing the visibility and reliability of innovative FCT security solutions.

Funded by
the European Union

REPORT FEEDBACK

We're collecting feedback on this report through the EU Survey Platform, if you'd like to share your thoughts anonymously please click on the link below.

https://ec.europa.eu/eusurvey/runner/enact-report-feedback

COPYRIGHT

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

This report was prepared by external experts outside of the ENACT consortium. The report does not represent the views or opinions of the ENACT project, and all content remains the sole responsibility of the authors.

USE OF AI

Following the ENACT policy for the use of AI, the author disclose that Generative AI tools were used in a limited supporting capacity for the creation of original illustrative figures and for the linguistic review and refinement of selected passages of this report. All substantive analysis, legal assessment, source and reference verification, conclusions and final editorial decisions were subject to human review and remained under the responsibility of the author.

ACRONYMS

AI	Artificial Intelligence
AI Act	Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence
AI RMF	Artificial Intelligence Risk Management Framework
ALTAI	Assessment List for Trustworthy Artificial Intelligence
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CNIL	Commission Nationale de l'Informatique et des Libertés
DPIA	Data Protection Impact Assessment
ECHR	European Convention on Human Rights
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
ELS	Ethical, Legal and Societal
ENACT	European Network Against Crime and Terrorism
EU	European Union
FCT	Fighting Crime and Terrorism
FRIA	Fundamental Rights Impact Assessment
GDPR	General Data Protection Regulation
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LEA	Law Enforcement Authority
LED	Law Enforcement Directive
NIST	National Institute of Standards and Technology
OECD	Organisation for Economic Co-operation and Development
OSINT	Open-Source Intelligence
R&I	Research and Innovation
RBOGM	Rights-Based Operational Governance Methodology
WP29	Article 29 Data Protection Working Party

06

EXECUTIVE SUMMARY

The increasing adoption of Artificial Intelligence (**AI**)-enabled intelligence-support and decision-support technologies is transforming the European Fight against Crime and Terrorism (**FCT**) landscape. AI-powered analytics, Open-Source Intelligence (**OSINT**), multilingual content analysis, visual analytics, graph-based intelligence and social media monitoring are significantly enhancing investigative capabilities, cross-border cooperation and situational awareness. At the same time, these technologies raise critical ethical, legal and societal (**ELS**) challenges concerning the protection of fundamental rights, proportionality, transparency, accountability, bias mitigation, human oversight and public trust. Ensuring that these technologies remain both operationally effective and aligned with European democratic values requires moving beyond regulatory compliance towards a governance model capable of demonstrating trustworthy AI deployment throughout the entire operational lifecycle.

The Advanced Report examines how rights-based governance can be operationalised for AI-enabled intelligence-support and decision-support technologies used by Law Enforcement Authorities (**LEAs**) in FCT contexts. The report focuses on translating legal, ethical and societal requirements into measurable and operational safeguards applicable throughout the AI lifecycle, from research and piloting to operational deployment. Particular attention is devoted to AI applications supporting intelligence analysis, OSINT and social media monitoring, multilingual content analysis, visual analytics and network intelligence. The analysis combines the applicable EU legal framework with recognised European and international AI governance and risk-management methodologies.

The report addresses three interrelated research questions. First, which ethical, legal and societal risks are most relevant when AI-enabled intelligence-support technologies assist criminal investigations, intelligence analysis and cross-border cooperation? Second, how can principles, regulatory requirements and governance measures – including legality, necessity, proportionality, non-discrimination, explainability, auditability, cybersecurity, data protection and meaningful human oversight – be translated into measurable operational safeguards capable of supporting trustworthy AI deployment?

Third, which governance mechanisms, implementation indicators and evidence-based validation approaches can assist LEAs, researchers and technology providers in demonstrating responsible AI deployment while ensuring compliance with European legal and ethical requirements?

To answer these questions, the report proposes a structured governance model organised around four complementary layers: (i) legal qualification and purpose limitation; (ii) fundamental rights and data protection impact assessment; (iii) AI governance, risk management and human oversight; and (iv) societal accountability through transparency, documentation, logging, redress mechanisms and independent review where appropriate. Building upon this framework, the report introduces **a Safeguards Effectiveness Matrix** that provides qualitative indicators, implementation criteria, and supporting evidence for assessing the maturity and effectiveness of each safeguard. Particular emphasis is placed on distinguishing technical risks (including bias, robustness, explainability, cybersecurity and data quality) from organisational and governance risks, such as unclear allocation of responsibilities, insufficient accountability, inadequate documentation, limited staff training, ineffective human oversight and the absence of continuous monitoring. The report argues that these organisational dimensions frequently represent the most significant obstacles to trustworthy operational deployment, even where technical solutions are available.

The proposed methodology combines regulatory analysis with evidence from representative operational scenarios that reflect realistic FCT use cases, enabling practical validation of the proposed governance model across diverse investigative contexts. The report also examines how safeguards can be monitored throughout the AI lifecycle using practical implementation indicators and governance evidence that support accountability, auditability and continuous improvement. By adopting a scenario-based and implementation-oriented perspective, the analysis aims to bridge the gap between regulatory obligations and day-to-day operational practice.

The outcome is a concise, policy-oriented and operationally relevant report that provides actionable recommendations for policymakers, LEAs, researchers and technology providers. The report explicitly complements existing ENACT analytical outputs by extending prior work from a governance implementation perspective. It also provides practical governance guidance, implementation checklists and recommendations to support the trustworthy adoption of AI-enabled intelligence technologies across the European security ecosystem. Ultimately, the report demonstrates that effective security, operational efficiency and the protection of fundamental rights should not be regarded as competing objectives but as mutually reinforcing conditions for lawful, accountable, trustworthy and socially legitimate AI deployment within European law enforcement.

1 - INTRODUCTION

08

CONTEXT, RELEVANCE AND SCOPE

Digital transformation has fundamentally reshaped the way information is collected, analysed and exploited within the European FCT ecosystem. Over the last decade, AI has progressively evolved from a research-oriented technology into an operational capability supporting intelligence analysis, criminal investigations and cross-border cooperation. Contemporary AI applications increasingly function as intelligence-support tools, assisting human operators in identifying relevant information, recognising patterns, prioritising investigative leads and processing large volumes of heterogeneous data originating from online environments, multilingual sources and complex digital ecosystems.

This transformation reflects a broader evolution in the way European law enforcement authorities approach information management. The exponential growth of publicly accessible digital information, the proliferation of social media platforms, encrypted communication channels, darknet marketplaces and transnational criminal networks have substantially increased the analytical burden placed upon investigators. AI-enabled technologies therefore represent an opportunity to improve investigative efficiency, situational awareness and operational coordination. At the same time, their growing integration within investigative workflows raises fundamental questions concerning the appropriate balance between technological innovation, operational effectiveness and the protection of fundamental rights.

Unlike systems designed to make or substantially automate decisions affecting individuals, most technologies considered in this report operate primarily as **intelligence-support and analytical-support systems**. They typically assist investigators by collecting, filtering, translating, classifying, correlating or visualising information without autonomously determining investigative or judicial outcomes. Nevertheless, their outputs (including classifications, alerts, confidence scores, relationship graphs, rankings or recommendations) may significantly influence subsequent human decisions. Consequently, the ethical, legal and societal implications of these technologies cannot be assessed solely by considering whether a final decision remains formally attributable to a human operator.

The rapid diffusion of AI-enabled investigative support has also exposed a widening gap between technological capabilities and governance mechanisms. Existing legal instruments provide an extensive framework governing the protection of personal data, fundamental rights and the deployment of AI systems.

However, significantly less attention has been devoted to the practical question of how these legal and ethical requirements should be translated into operational safeguards capable of guiding everyday investigative practice. In other words, while the applicable regulatory obligations are increasingly well defined, the mechanisms through which compliance, accountability and trustworthiness can be demonstrated throughout the operational lifecycle remain comparatively underdeveloped.

The European regulatory landscape has recently evolved through the adoption of **Regulation (EU) 2024/1689** (hereinafter also **"AI Act"**), which complements existing data protection and fundamental rights instruments by introducing a harmonised risk-based governance framework for AI systems. As stated in Recital 1, the AI Act seeks to promote the uptake of *"human-centric and trustworthy artificial intelligence"* while ensuring *"a high level of protection of health, safety and fundamental rights"*.[1] This framework complements the existing guarantees established by the Charter of Fundamental Rights of the European Union (particularly Articles 1, 7, 8, 11, 20, 21, 47 and 48)[2] and by the European Convention on Human Rights (notably Articles 6, 8, 10, 13 and 14),[3] which collectively require that the deployment of security technologies remains compatible with the principles of legality, necessity, proportionality, non-discrimination and effective judicial protection.

Within this evolving legal environment, governance cannot be reduced to regulatory compliance alone. Compliance represents a necessary condition for the lawful deployment of AI-enabled technologies, but it does not automatically ensure that these technologies are ethically justified, socially legitimate or operationally trustworthy. As recognised by the European Data Protection Supervisor (**EDPS**) in its checklist on human intervention (hereinafter also "**EDPS Checklist on Human Intervention**"), meaningful human intervention requires considerably more than the formal presence of a human decision-maker; it depends upon governance arrangements capable of ensuring effective oversight, accountability, transparency and genuine human control throughout the decision-making process.[4] Likewise, Europol's ethical decision-making methodology (hereinafter also **"Europol Methodology"**) emphasises that technological choices should be evaluated not only against legal requirements, but also against the values underpinning democratic societies, recognising that ethical analysis becomes particularly relevant where technological innovation develops more rapidly than the applicable regulatory framework.[5]

Against this background, the present report adopts a focused scope. It concentrates on **AI-enabled intelligence-support technologies** that assist law enforcement authorities during information gathering, intelligence analysis, and investigative activities. Particular attention is devoted to OSINT, social media monitoring, multilingual and multimodal content analysis, graph-based intelligence and related analytical technologies. These use cases were selected because they represent some of the most mature and operationally relevant AI applications currently supporting European law enforcement activities while simultaneously presenting significant ethical, legal and societal challenges.

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Recital 1. Regulation - EU - 2024/1689 - EN - EUR-Lex.

[2] Charter of Fundamental Rights of the European Union, Articles 1, 7, 8, 11, 20, 21, 47 and 48. EUR-Lex - 12012P/TXT - EN - EUR-Lex.

[3] European Convention on Human Rights, Articles 6, 8, 10, 13 and 14. European Convention on Human Rights.

[4] European Data Protection Supervisor (EDPS), "Checklist on human intervention on automated decision-making (ADM)". Checklist on Human oversight of automated decision-making.

[5] Europol Innovation Lab, "Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making", revised edition, January 2025. Assessing technologies in law enforcement: a method for ethical decision-making | Europol.

10

The report also seeks to complement existing analytical work developed within the ENACT project and by Europol, shifting the focus from **what requirements apply** to **how those requirements can be operationalised, evidenced and continuously monitored** through practical governance mechanisms, measurable safeguards and scenario-based validation. In doing so, it aims to contribute to a more mature understanding of trustworthy AI governance within the European FCT research and innovation landscape.

OBJECTIVES, RESEARCH QUESTIONS AND STRUCTURE

This report develops a structured analytical framework aimed at operationalising rights-based governance for AI-enabled intelligence-support technologies throughout their lifecycle, from research and experimentation to operational deployment. The underlying assumption is that trustworthy AI cannot be demonstrated merely through compliance with individual legal obligations or the adoption of high-level ethical principles. Instead, it requires an integrated governance approach capable of translating legal, ethical and societal requirements into concrete safeguards, supported by objective evidence and subject to continuous monitoring and review.

The originality of the proposed approach lies in combining several complementary assessment perspectives that have traditionally been examined separately. In fact, the report proposes an integrated governance model capable of linking these dimensions within a common analytical framework. Particular emphasis is placed on distinguishing **technical risks** – such as data quality issues, algorithmic bias, robustness limitations or cybersecurity vulnerabilities – from **organisational risks**, including inadequate governance structures, insufficient documentation, ineffective oversight, lack of training, automation bias, and unclear allocation of responsibilities. The report argues that these organisational dimensions frequently represent the principal barriers to trustworthy operational deployment, even where the underlying technological solutions demonstrate satisfactory technical performance.

To operationalise this approach, the report introduces a **Safeguards Effectiveness Matrix**, designed to support the systematic identification, implementation and monitoring of governance measures throughout the AI lifecycle. The matrix builds upon the structured logic commonly adopted in Data Protection Impact Assessments (**DPIAs**),[6] ethical impact assessments and AI governance frameworks, distinguishing between implemented safeguards, additional measures, and the expected evolution of residual risk following their implementation.[7] It further incorporates indicators derived from the EDPS Checklist on Human Intervention, recognising that effective human oversight requires demonstrable organisational arrangements, documented procedures, appropriate training, decision logging, auditability and continuous review, rather than the merely formal inclusion of a human actor within an automated workflow.

[6] Regulation (EU) 2016/679, Article 35; Directive (EU) 2016/680, Article 27; Article 29 Working Party, Guidelines on Data Protection Impact Assessment (WP248 rev.01); European Data Protection Board, Opinion 12/2018

[7] CNIL, Self-assessment Guide for AI Systems. Self-assessment guide for artificial intelligence (AI) systems | CNIL.
High-Level Expert Group on AI, Assessment List for Trustworthy Artificial Intelligence (ALTAI). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment | Shaping Europe's digital future.

The remainder of the report is organised as follows.

Section 2 describes the analytical methodology, the legal and policy sources underpinning the assessment and the scenario-based approach adopted throughout the analysis.

Section 3 examines the operational context of AI-enabled intelligence-support technologies and identifies the principal ethical, legal and societal challenges arising from their deployment.

Section 4 develops the proposed rights-based governance framework by analysing the applicable legal regime, impact assessment methodologies and ethical decision-making processes.

Section 5 introduces the four-layer safeguards model together with the Safeguards Effectiveness Matrix and proposes measurable implementation indicators.

Section 6 applies the framework to representative operational scenarios and discusses its implications for the European FCT research and innovation ecosystem.

Section 7 formulates recommendations for policymakers, law enforcement authorities, researchers and technology providers.

Finally, Section 8 summarises the principal conclusions and identifies future directions for the operational governance of AI-enabled intelligence-support technologies.

2- METHODOLOGY AND EVIDENCE BASE

ANALYTICAL METHODOLOGY

The analytical approach adopted in this report is based on the premise that the governance of AI-enabled intelligence-support technologies cannot be adequately assessed through a single disciplinary perspective. Technologies supporting intelligence analysis, OSINT, multilingual content processing, graph-based intelligence, and other analytical functions operate within complex socio-technical environments where legal compliance, ethical acceptability, organisational governance, and technical performance are closely interconnected. Assessing only one of these dimensions would inevitably provide an incomplete picture of the conditions required for trustworthy deployment.

For this reason, the report adopts an integrated **Rights-Based Operational Governance Methodology (RBOGM)**, combining legal analysis, ethical assessment, scenario-based analysis, impact assessment, safeguard mapping, and evidence-based validation within a single analytical framework. The methodology examines the broader governance ecosystem in which AI-enabled intelligence-support technologies are developed, deployed, and used. This reflects the growing recognition that many of the risks associated with AI systems arise not solely from algorithmic behaviour, but from organisational arrangements, operational practices, human decision-making, and institutional accountability.[8]

The proposed methodology is grounded in the European fundamental-rights framework and follows a risk-based, lifecycle-oriented approach. It builds upon established methodologies developed in the fields of data protection, trustworthy AI, ethical impact assessment, and law enforcement innovation, while adapting them to the specific characteristics of AI-enabled intelligence-support technologies operating in the FCT domain. In particular, the methodological design draws inspiration from the integrated assessment approach adopted in recent European research activities, where legal, ethical, technical and organisational dimensions are evaluated through structured evidence, documented safeguards, and progressive risk reassessment, rather than through a purely compliance-oriented exercise.[9]

[8] European Commission, Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, 2019. Ethics Guidelines for AI.

[9] The methodological structure builds upon multidisciplinary assessment approaches combining legal, ethical, technical and organisational analysis, including Regulation (EU) 2016/679, Article 35; Directive (EU) 2016/680, Article 27; Article 29 Working Party, Guidelines on Data Protection Impact Assessment (WP248 rev.01); European Data Protection Board, Opinion 12/2018; CNIL, "Self-assessment Guide for AI Systems"; High-Level Expert Group on AI, "Assessment List for Trustworthy Artificial Intelligence (ALTAI)"; Europol Innovation Lab, "Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making", revised edition, January 2025.

Unlike conventional compliance assessments, which frequently focus on determining whether predefined legal obligations have been formally satisfied, the methodology proposed here seeks to answer a broader question: **under which conditions can an AI-enabled intelligence-support technology be considered operationally trustworthy throughout its lifecycle?** Addressing this question requires moving beyond the identification of applicable legal provisions towards the systematic evaluation of how safeguards are implemented, monitored, evidenced, and periodically reviewed within operational environments.

To achieve this objective, the methodology combines seven complementary analytical stages that are applied consistently throughout the report.

The first stage consists of the **legal qualification of the technology and its operational context**. Before any assessment of risks or safeguards can be undertaken, it is necessary to identify the intended purpose of the system, the operational environment in which it is deployed, the actors involved, the categories of data processed and the applicable legal regime. Particular attention is paid to distinguishing between research, testing, operational deployment contexts, and to identifying the respective responsibilities of providers, deployers, competent authorities, and other relevant actors. This preliminary qualification also provides the basis for determining the relevance of the applicable data protection framework and, where appropriate, the AI Act risk-based classification.[10]

The second stage concerns the **identification of the rights, interests and legitimate expectations potentially affected by the deployment of the technology**. The methodology adopts a broader fundamental-rights perspective encompassing equality, non-discrimination, freedom of expression, procedural fairness, effective remedies, human dignity, and the protection of vulnerable groups. This stage also considers the interests of law enforcement operators, technology providers, and society at large, recognising that trustworthy AI governance requires balancing multiple legitimate objectives rather than maximising a single policy goal.

The third stage consists of the **identification and categorisation of risks** analysed according to three complementary dimensions.

1. The first comprises **technical risks**, including deficiencies in data quality, robustness, explainability, cybersecurity, model reliability, performance degradation, and algorithmic bias.
2. The second concerns **organisational and governance risks**, such as inadequate documentation, insufficient allocation of responsibilities, weak oversight mechanisms, lack of operator training, ineffective audit procedures, automation bias, and deficiencies in change management.
3. The third dimension encompasses **ethical and societal risks**, including disproportionate surveillance, indirect discrimination, function creep, chilling effects, reduced institutional legitimacy, and adverse impacts on victims, vulnerable individuals, or communities.

This tripartite categorisation reflects the understanding that trustworthy deployment depends as much upon governance arrangements as upon technological performance.

[10] Regulation (EU) 2024/1689, Articles 3 and 6, Annex III; European Commission, "Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)". The Commission publishes guidelines on AI system definition to facilitate the first AI Act's rules application | Shaping Europe's digital future.

The fourth stage focuses on the **identification and analysis of safeguards**. The methodology adopts a multidimensional perspective encompassing legal, organisational, procedural and technical measures. Safeguards are therefore examined in relation to purpose limitation, governance arrangements, human oversight, documentation, transparency, auditability, cybersecurity, accountability, training, incident management, and continuous monitoring. Particular attention is devoted to identifying the actor responsible for implementing each safeguard and to distinguishing safeguards embedded by technology providers from those requiring operational implementation by deployers or competent authorities.

The fifth stage concerns the **identification of supporting evidence**. A safeguard cannot be regarded as effective merely because it has been described in policy documentation or system specifications. Instead, the methodology requires the existence of objective evidence capable of demonstrating its actual implementation. Such evidence may include documented procedures, governance policies, audit logs, technical testing results, validation reports, operator training records, monitoring mechanisms, performance indicators, incident registers or independent assessment outcomes. This evidence-based perspective represents one of the distinguishing features of the proposed analytical framework, shifting the focus from declaratory compliance towards demonstrable operational assurance.

The sixth stage consists of the **evaluation of safeguard effectiveness and residual risk**. Inspired by established impact-assessment methodologies, the analysis distinguishes between safeguards that are already implemented and evidenced at the time of assessment and additional measures that are recommended for future implementation. Consequently, the methodology differentiates between the **current governance position**, reflecting the safeguards demonstrably in place, and the **projected governance position**, representing the expected residual risk following the implementation of identified improvements. This distinction is essential to avoid conflating existing controls with planned measures and to ensure that recommendations remain transparent regarding their implementation status.

The seventh stage translates the analytical findings into **practical governance recommendations**. The methodology seeks to formulate recommendations that are operationally actionable, proportionate and attributable to specific categories of stakeholders, including policymakers, law enforcement authorities, researchers, technology developers, procurement authorities and oversight bodies. The objective is not to prescribe a single governance model but to provide a structured framework capable of supporting informed decision-making across different operational contexts.

The methodology is further strengthened through a **scenario-based analytical approach**. Instead of assessing AI technologies in isolation, the report examines representative operational scenarios reflecting realistic intelligence-support activities within the FCT domain. This approach recognises that identical technologies may generate substantially different ethical, legal and societal implications depending on their intended purpose, operational environment, users, affected persons, and downstream reliance. Scenarios therefore serve not merely as illustrative examples, but as analytical instruments enabling the contextual evaluation of risks, safeguards, and governance measures.

Finally, the methodology incorporates an iterative **validation logic.** Governance should not be understood as a one-off compliance exercise conducted before deployment but as a continuous process requiring periodic reassessment throughout the AI lifecycle. New operational environments, updated datasets, evolving investigative practices, changes in legislation or modifications to system functionalities may all alter the overall risk profile of a technology. Consequently, safeguards, supporting evidence, and governance arrangements should be reviewed whenever material changes occur, ensuring that the deployment of AI-enabled intelligence-support technologies remains lawful, ethically justified, operationally effective, and aligned with fundamental rights over time.

NORMATIVE, POLICY AND ACADEMIC SOURCES

The analytical framework presented in this report is supported by a structured hierarchy of legal, policy, technical and academic sources. Given the multidisciplinary nature of AI-enabled intelligence-support technologies, no single legal instrument or governance framework is capable of addressing all ethical, legal, organisational and technical issues arising throughout the technology lifecycle. Consequently, the analysis adopts a layered approach in which different categories of sources perform complementary functions within the overall assessment methodology.

Binding legal instruments establish the legal obligations governing the deployment of AI-enabled technologies and define the minimum level of protection that should be guaranteed to individuals. Soft law and institutional guidance provide interpretative criteria, operational methodologies, and best practices supporting the practical implementation of those legal obligations. International standards contribute with structured governance and assurance mechanisms capable of strengthening organisational maturity and demonstrating accountability. Finally, academic literature provides the conceptual and methodological foundations necessary to interpret emerging challenges that have not yet been fully addressed by legislation or regulatory guidance.

Throughout this report, these different sources are therefore considered according to their respective normative value and analytical function, avoiding any equivalence between legally binding rules and voluntary governance instruments.

Binding legal sources. The primary legal framework underpinning the analysis consists of the Charter of Fundamental Rights of the European Union (the **"Charter"**), the European Convention on Human Rights (the **"ECHR"**), Regulation (EU) 2016/679 (**General Data Protection Regulation – "GDPR"**), Directive (EU) 2016/680 (**Law Enforcement Directive – "LED"**), Regulation (EU) 2024/1689 (**AI Act**), together with applicable national legislation governing criminal procedure, law enforcement activities, and the implementation of Union law. These instruments constitute the legal foundation upon which the proposed governance model is built.

The Charter provides the principal constitutional framework governing the protection of fundamental rights within the European Union, including human dignity, privacy, data protection, equality, non-discrimination, freedom of expression, effective judicial protection, and the presumption of innocence.

Within this broader constitutional framework, the GDPR and the LED establish complementary, but distinct, legal regimes governing the processing of personal data. Their applicability depends not on the technology itself, but on the nature of the controller, the purpose of processing and the legal context in which processing takes place. Accordingly, the report avoids assuming that all AI-enabled investigative technologies are subject to identical legal obligations, instead recognising that the applicable regime must always be determined through a case-specific legal qualification.

The AI Act complements this existing framework by introducing a harmonised risk-based governance regime for AI systems. This Regulation establishes additional obligations concerning the design, development, deployment, and oversight of AI systems according to their intended purpose and risk profile. The report therefore treats the AI Act as an integral component of the broader European governance architecture, analysing its interaction with pre-existing legal obligations.

Soft law and institutional guidance. Binding legislation is complemented by a substantial body of institutional guidance that supports its practical interpretation and implementation. Although these instruments do not generally create legally enforceable obligations, they play a crucial role in promoting consistent interpretation, operational convergence and good administrative practice across EU Member States.

With regard to data protection, particular importance is attached to the **Article 29 Working Party Guidelines on Data Protection Impact Assessment (WP248 rev.01)**,[11] complemented by the **European Data Protection Board (EDPB) Data Protection Impact Assessment (DPIA) Template**.[12] Together, these instruments provide both the methodological principles and the practical structure for identifying high-risk processing operations, assessing risks to the rights and freedoms of natural persons, documenting mitigation measures, and demonstrating accountability throughout the lifecycle of processing activities. Their contribution extends beyond the formal conduct of a DPIA by promoting evidence-based assessment, proportionality, continuous review, and accountability as core governance principles.

The report also relies upon the **EDPS Checklist on Human Intervention** concerning meaningful human intervention and AI-supported decision-making. The EDPS adopts a broader organisational perspective, emphasising competence, authority, information availability, workload, governance arrangements, and effective intervention capabilities as essential conditions for meaningful oversight. These principles are particularly relevant for intelligence-support technologies, where the quality of human interaction frequently determines whether AI outputs are critically assessed or accepted without sufficient scrutiny.

Ethical governance is also informed by the **Europol methodology *"Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making,"*** developed by the Europol Innovation Lab together with the European Clearing Board and revised in 2025. This methodology introduces a structured value-based decision-making process requiring decision-makers to identify the relevant facts, affected stakeholders, competing values, available options and foreseeable consequences before selecting and justifying a particular technological solution. The methodology therefore complements legal analysis by addressing ethical dilemmas that may arise even where no explicit legal prohibition exists.

The report additionally considers the Council of Europe Guidelines on Artificial Intelligence and Data Protection, which advocate a human rights-based approach centred on dignity, democratic values, impact assessment, and preventive risk management throughout the lifecycle of AI systems.[13] Their relevance lies in reinforcing the principle that technological innovation should remain subordinate to the protection of fundamental rights rather than redefining their scope.

[11] Article 29 Working Party, "Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679 (WP248 rev.01)", as endorsed by the European Data Protection Board. ARTICLE29 - Guidelines on Data Protection Impact Assessment (DPIA) (wp248rev.01).

[12] European Data Protection Board, Data Protection Impact Assessment (DPIA) Template, 2026. Template for Data Protection Impact Assessment | European Data Protection Board

[13] Council of Europe, "Guidelines on Artificial Intelligence and Data Protection", T-PD(2019)01.

Broader governance perspectives are provided by the **OECD AI Principles**,[14] the **Assessment List for Trustworthy Artificial Intelligence (ALTAI)** developed by the High-Level Expert Group on AI, and the **NIST Artificial Intelligence Risk Management Framework (AI RMF)**.[15] Although differing in scope and institutional origin, these frameworks converge around several core concepts, including accountability, transparency, human oversight, robustness, continuous monitoring, and lifecycle risk management. Throughout this report, they are treated as governance frameworks supporting organisational maturity.

Standards and technical governance frameworks. The analysis further incorporates internationally recognised standards addressing organisational governance, risk management, and information security. Particular attention is devoted to **ISO/IEC 42001**,[16] the first international management system standard specifically dedicated to artificial intelligence governance, and **ISO/IEC 23894**,[17] which provides guidance for AI risk management across the system lifecycle. These standards offer structured approaches for assigning responsibilities, documenting governance processes, managing organisational risks and supporting continuous improvement. Their contribution lies primarily in operationalising governance practices rather than determining the legality of specific AI applications.

Where relevant, reference is also made to the **ISO/IEC 27000 family**, particularly those standards concerning information security management, risk assessment, security controls, and incident management, as well as to emerging **CEN-CENELEC standardisation activities** supporting the implementation of the AI Act. These technical standards are considered valuable assurance mechanisms capable of strengthening governance maturity, facilitating internal audits and supporting conformity assessment processes. However, they are not treated as equivalent to legislation and cannot substitute compliance with applicable legal obligations. Conformity with recognised standards may support the demonstration of good governance practices, but it does not in itself establish the lawfulness or ethical acceptability of a particular technological deployment.

Academic doctrine. Finally, the report is informed by a growing body of interdisciplinary academic literature addressing the governance of AI within law enforcement, public administration and high-impact decision-making.

Academic contributions are particularly important in areas where legislative and regulatory frameworks continue to evolve. In relation to **fundamental rights impact assessment**, the work of **Alessandro Mantelero** has significantly contributed to conceptualising the FRIA as an assessment methodology extending beyond traditional data protection analysis by systematically examining the broader societal and collective impacts of emerging technologies.[18] This perspective is complemented by the **ALIGNER Fundamental Rights Impact Assessment methodology**, which adapts rights-based assessment to AI applications within law enforcement and security contexts through structured stakeholder analysis, contextual evaluation and multidisciplinary risk identification.[19]

[14] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2019.
[15] NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).
[6] ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. ISO/IEC 42001:2023 - AI management systems.
[17] ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management. ISO/IEC 23894:2023 - AI — Guidance on risk management.
[18] Mantelero, A., "The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template", 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template by Alessandro Mantelero :: SSRN.
[19] ALIGNER Project, Fundamental Rights Impact Assessment Methodology for AI in Law Enforcement. Fundamental Rights Impact Assessment (FRIA) | aligner.

The report also draws upon the work of **Sterz et al.**, whose interdisciplinary analysis demonstrates that meaningful human oversight depends upon the interaction of legal, organisational, cognitive and technological factors, rather than the mere existence of human intervention within a decision-making process.[20] This conclusion aligns closely with the governance-oriented perspective adopted throughout the present analysis.

With regards to algorithmic fairness, the report considers the influential work of **Wachter, Mittelstadt, and Russell**, who highlight the conceptual gap between statistical notions of algorithmic fairness and the legal requirements governing discrimination under European law.[21] Their analysis illustrates why technical fairness metrics cannot automatically demonstrate compliance with non-discrimination principles and why legal assessment remains indispensable.

Finally, recent research concerning **explainability, human–AI interaction, and appropriate reliance** is incorporated to address one of the central governance challenges associated with intelligence-support technologies. These studies demonstrate that increasing the explainability of AI outputs does not necessarily improve decision quality or reduce automation bias. Effective governance therefore requires not only explainable systems, but also organisational arrangements enabling operators to critically interpret AI outputs, recognise uncertainty, and exercise independent professional judgement before relying upon algorithmic recommendations.[19]

Taken together, these legal, institutional, technical and academic sources form the multidisciplinary evidence base underpinning the analytical framework developed throughout this report. Their combined use reflects the central premise advanced in the previous sections: trustworthy AI governance cannot be derived from compliance with a single regulatory instrument, but requires the coherent integration of legal obligations, ethical principles, organisational governance, technical assurance mechanisms, and evidence-based operational practice.

[20] Sterz, A. et al., "On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives", 2024. On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives | Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency.

[21] Wachter, S., Mittelstadt, B., & Russell, C., "Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI", 2024. Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI by Sandra Wachter, Brent Mittelstadt, Chris Russell :: SSRN.

3- OPERATIONAL CONTEXT AND ELS CHALLENGES

AI-ENABLED INTELLIGENCE SUPPORT IN PRACTICE

Artificial intelligence is increasingly embedded within the information lifecycle, supporting modern law enforcement activities. Instead of replacing investigators or autonomously determining operational decisions, AI-enabled systems primarily function as **intelligence-support technologies**, assisting human operators in processing information that would otherwise exceed the practical limits of manual analysis. Their principal contribution lies in reducing the time required to identify relevant information, improving the consistency of analytical processes and enabling investigators to extract meaningful insights from heterogeneous and continuously evolving digital environments.

From an operational perspective, intelligence support can be understood as a sequence of interconnected analytical activities through which large volumes of information are progressively transformed into actionable knowledge. Although individual technologies differ considerably in terms of architecture, underlying algorithms and operational maturity, they generally contribute to one or more stages of a common investigative workflow, beginning with information acquisition and concluding with human interpretation and operational decision-making.

The first stage concerns **information collection and filtering**. Contemporary investigations frequently involve the analysis of information originating from publicly accessible websites, social media platforms, online forums, darknet marketplaces, open databases, financial records, or digital evidence collected during criminal investigations. The quantity of potentially relevant information often exceeds the analytical capacity of investigators, making automated filtering and prioritisation essential. AI-enabled systems support this activity by identifying potentially relevant content according to predefined investigative parameters, reducing duplication, highlighting emerging topics and facilitating the organisation of information before substantive analysis begins. Rather than replacing investigative judgement, these technologies assist operators in identifying where further human attention should be directed.

Following collection, information frequently requires **translation, normalisation and content analysis**. Criminal investigations increasingly involve multilingual communications, informal language, abbreviations, slang, coded terminology, and multimedia content combining text, images, audio, and video.

AI-supported language technologies facilitate the translation of multilingual material, identify relevant concepts, and classify documents according to their potential investigative relevance. At the same time, multimodal analytical capabilities allow investigators to examine different forms of digital evidence through an integrated perspective, reducing the need to analyse each information source independently. These functionalities contribute significantly to cross-border cooperation, where linguistic diversity would otherwise constitute a substantial operational obstacle.

A further analytical step consists of **entity and relationship extraction**. Investigators often need to identify recurring persons, organisations, locations, events, financial accounts, vehicles or digital identifiers appearing across multiple sources. AI-enabled analytical tools can automatically detect these entities and identify relationships among them, generating structured representations of information that would otherwise remain dispersed throughout large datasets. Relationship extraction similarly assists investigators in identifying associations between entities, enabling the progressive reconstruction of networks, communication patterns and potential criminal structures while reducing the manual effort traditionally required for information correlation.

Structured information can subsequently be explored through **graph analysis and network intelligence**. Graph-based approaches allow investigators to visualise complex relationships among individuals, organisations, digital identities, transactions or communications, facilitating the identification of central actors, intermediary nodes, previously unknown connections or structural characteristics of criminal networks. Therefore, graph analytics can support exploratory reasoning by highlighting relationships that may warrant further verification through conventional investigative methods. Consequently, graph-based intelligence functions primarily as an analytical aid rather than an autonomous evidential instrument.

AI-enabled intelligence support also increasingly incorporates **anomaly detection** capabilities. Instead of searching exclusively for predefined patterns, these techniques assist investigators in identifying behaviours, transactions or communication activities that deviate from expected or historically observed patterns. Such anomalies may indicate previously unknown criminal methodologies, emerging trafficking routes, coordinated online activities or suspicious financial movements requiring further investigation. However, anomalous behaviour does not necessarily imply unlawful conduct. The analytical value of anomaly detection therefore depends upon subsequent human assessment capable of distinguishing genuinely relevant signals from statistically unusual but entirely legitimate activities.

Throughout the analytical process, AI systems frequently support investigators by generating **rankings, confidence indicators, alerts and visual representations** intended to facilitate prioritisation and situational awareness. Ranking mechanisms may identify information considered potentially relevant according to predefined analytical criteria, while alerts notify operators of emerging patterns or significant changes requiring attention. Visual dashboards, timelines, geospatial representations, and network graphs further assist investigators in understanding complex datasets by presenting analytical results in forms that are more readily interpretable than unstructured information alone. These functionalities contribute to operational efficiency but also influence how investigators allocate attention, interpret evidence, and prioritise investigative activities.

The final stage remains **human interpretation and investigative reasoning**. AI-enabled intelligence-support technologies **do not eliminate the need for professional judgement**; rather, they reshape the manner in which investigators interact with information. Operators remain responsible for assessing the relevance, reliability and evidential value of analytical outputs, integrating AI-generated information with contextual knowledge, investigative experience, and additional sources of evidence.

In practice, AI-generated outputs frequently constitute one element within a broader evidential and analytical process, supporting – but not replacing – the exercise of professional discretion.

This operational perspective highlights an important characteristic shared by most intelligence-support technologies considered throughout this report. Their principal impact does not arise from fully automated decision-making, but from their capacity to influence how information is selected, interpreted and acted upon by human operators. A recommendation, relationship graph, anomaly alert or confidence score may substantially shape subsequent investigative reasoning even where no automated decision is formally adopted. Consequently, the governance challenges associated with these technologies extend beyond algorithmic accuracy and encompass broader questions concerning transparency, explainability, data provenance, human oversight, organisational accountability, and appropriate reliance on AI-supported analysis.

The transformation of information into operational knowledge should therefore be understood not as a purely technological process but as a **human–AI collaborative workflow**, in which each analytical stage introduces specific opportunities and corresponding governance challenges. Information collected from multiple digital environments must remain relevant, reliable and lawfully obtained; analytical outputs must be interpreted within their operational context; and human operators must retain both the capacity and the responsibility to critically assess AI-generated recommendations before they influence investigative decisions. This perspective provides the foundation for the following sections, which examine how representative operational scenarios give rise to different categories of ethical, legal, and societal risks and why those risks require governance mechanisms extending beyond technical system performance alone.

CROSS-CUTTING RISKS: TECHNICAL, ORGANISATIONAL AND SOCIETAL

The representative scenarios examined in the previous section demonstrate that the risks associated with AI-enabled intelligence-support technologies rarely originate from a single source. Instead, they emerge through the interaction between technical characteristics, organisational practices and the operational environment in which AI outputs are interpreted and used. Consequently, the governance of AI systems should not focus exclusively on algorithmic performance or legal compliance in isolation but should consider how risks propagate across the entire socio-technical system throughout the AI lifecycle.[22]

This systemic perspective is increasingly reflected in European AI governance frameworks, including the AI Act, particularly its provisions on risk management, human oversight and technical robustness, as well as the High-Level Expert Group's Ethics Guidelines for Trustworthy AI and the Assessment List for Trustworthy Artificial Intelligence, which recognise that technical reliability, organisational accountability and the protection of fundamental rights represent complementary (but conceptually distinct) dimensions of trustworthy AI.[23]

[21] NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, emphasising that AI risks emerge from interactions among technical systems, human actors, organisational processes and operational contexts; ISO/IEC 23894:2023, *Artificial Intelligence — Guidance on Risk Management*.

[23] See Regulation (EU) 2024/1689, in particular Recitals 1 and 27 and Articles 9, 14 and 15; European Commission, High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (2019); High-Level Expert Group on Artificial Intelligence, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*, 2020.

A system may exhibit satisfactory technical performance while nevertheless producing unlawful, disproportionate or socially harmful outcomes if deployed for purposes beyond its intended scope, operated without effective oversight or embedded within inadequate governance arrangements. Conversely, robust organisational procedures cannot fully compensate for unreliable analytical outputs generated by poor-quality data or technically deficient models. Trustworthiness and fundamental-rights compliance should therefore be assessed jointly, while recognising that they address different categories of risk.[24]

The first category concerns **technical risks**, which affect the reliability and quality of AI-generated outputs. These include inaccurate, incomplete or outdated datasets, weak data provenance, false positives and negatives, model and data drift, insufficient robustness against changing operational conditions, cybersecurity vulnerabilities, limited explainability, and the risk of re-identification where multiple datasets are combined. Individually, these issues may appear primarily technical; however, their operational significance lies in the fact that they directly influence the quality of the analytical information presented to investigators. Unreliable or insufficiently contextualised outputs may subsequently shape investigative priorities, resource allocation or operational decision-making in ways that are not immediately apparent to human operators.

Technical limitations frequently give rise to a second layer of **organisational risks**, reflecting how AI-enabled technologies are governed, integrated and supervised within law enforcement organisations. Even technically reliable systems may produce problematic outcomes where responsibilities are poorly defined, operators receive insufficient training, escalation procedures are absent, workloads encourage excessive reliance on automated outputs, or organisational processes fail to ensure adequate auditability and continuous monitoring. Additional challenges arise from excessive dependence on technology providers, inadequate change management following system updates, and the gradual use of AI systems beyond their originally intended purpose. These risks are particularly significant because they affect the institutional capacity to critically assess AI-generated outputs rather than the outputs themselves. As emphasised by the EDPS, meaningful human oversight depends not simply on retaining a human "in the loop", but on ensuring that operators possess the competence, authority, information and organisational conditions necessary to effectively challenge or disregard automated recommendations where appropriate.[25]

The interaction between technical and organisational deficiencies may ultimately generate broader societal and fundamental-rights risks. Poor data quality combined with insufficient human oversight may contribute to disproportionate surveillance practices; inaccurate relationship detection may reinforce indirect discrimination or create unjustified suspicion through guilt by association; excessive reliance on opaque analytical outputs may undermine the presumption of innocence, reduce opportunities for effective challenge and weaken procedural fairness.

[24] High-Level Expert Group on Artificial Intelligence, *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*; Regulation (EU) 2024/1689 (AI Act), Recitals 1, 27 and relevant provisions concerning risk management and human oversight; OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449).

[25] European Data Protection Supervisor, "*Checklist on Human Intervention in AI-supported Decision-Making*"; EDPS, TechDispatch: Human Oversight. See also Sterz, A. et al., "*Meaningful Human Oversight of AI Systems: An Interdisciplinary Perspective*", demonstrating that effective oversight depends upon organisational, legal, cognitive and technical conditions rather than the mere presence of a human operator within the decision-making process.

At a broader institutional level, these dynamics may also erode public confidence in law enforcement authorities, particularly where individuals are unable to understand how AI-supported conclusions were reached or where governance mechanisms fail to provide transparency, accountability and effective remedies. Vulnerable groups (including children, minorities, journalists, human rights defenders, and persons incidentally captured during intelligence collection) may experience these impacts disproportionately, further amplifying existing structural inequalities.

Therefore, it is more appropriate to understand these risks as **interdependent stages within a broader chain of risk propagation**. Weak data provenance may reduce the reliability of analytical outputs; unreliable outputs may increase the likelihood of automation bias and inappropriate reliance by investigators; inappropriate reliance may distort investigative prioritisation or resource allocation; these operational decisions may in turn result in disproportionate surveillance, discriminatory outcomes or unjustified investigative attention directed towards particular individuals or communities. Ultimately, repeated governance failures of this nature may undermine institutional legitimacy and reduce public trust in the lawful use of AI-enabled investigative technologies.

*Figure 1 : Illustrative risk propagation across the AI-supported investigative lifecycle**

This cascading perspective has important implications for governance. It suggests that effective safeguards should not be designed solely to mitigate isolated technical failures but should interrupt the propagation of risk at multiple stages of the AI lifecycle. Improving data quality, documenting provenance, strengthening human oversight, clarifying organisational responsibilities, ensuring continuous monitoring, and maintaining effective accountability mechanisms are therefore not independent mitigation measures, but complementary governance interventions addressing different points within the same causal chain.

* Author's own elaboration, informed by the risk categories and governance framework discussed in this report.

4 - LEGAL AND ETHICAL GOVERNANCE FRAMEWORK

LEGAL QUALIFICATION, APPLICABLE REGIME AND ACCOUNTABILITY

The governance of AI-enabled intelligence-support technologies begins long before questions concerning algorithmic performance or technical safeguards arise. The first and arguably most important step consists of determining the legal framework governing the intended deployment of the technology. This preliminary qualification is not merely a formal legal exercise, but establishes the obligations, responsibilities, and accountability mechanisms that will apply throughout the system lifecycle. Incorrectly identifying the applicable legal regime may undermine subsequent impact assessments, invalidate governance measures, and ultimately compromise both operational effectiveness and the protection of fundamental rights.

A recurring misconception in discussions concerning AI for law enforcement is the assumption that the legal framework automatically follows the nature of the technology itself. European law adopts a fundamentally different approach. Neither the GDPR, the LED, nor the AI Act applies solely because artificial intelligence is used. Instead, their applicability depends upon a combination of factors, including the intended purpose of the processing, the identity of the actor determining the purposes and means of processing, the operational context in which the system is deployed, and the role that AI-generated outputs play within subsequent decision-making processes.

Accordingly, the legal qualification of an AI-enabled intelligence-support system should always begin with a structured assessment of its **intended purpose**. Under the AI Act, intended purpose constitutes the reference point for determining how a system is designed to operate, the functions it performs and the foreseeable context in which its outputs will be used. The same functionality may therefore generate different legal obligations depending upon whether it is deployed for research purposes, intelligence support, operational investigations, administrative activities or strategic analysis. Likewise, an identical technological capability may remain outside the scope of high-risk AI regulation in one context while becoming subject to additional obligations in another because its outputs influence decisions affecting natural persons.[26]

[26] Regulation (EU) 2024/1689, Articles 3 and 6, Annex III; European Commission, "*Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI ACT)*".

This contextual approach also governs the relationship between the GDPR and the LED. These two legal instruments should not be regarded as simultaneously governing the same processing operation. Instead, they establish **distinct legal regimes** that apply according to the specific purpose of processing and the status of the entity determining the purposes and means of that processing. The GDPR generally governs processing carried out outside the scope of criminal law enforcement activities, including many research, development, testing and validation activities undertaken by universities, research organisations, technology providers or other entities participating in innovation projects. Conversely, the LED applies where personal data are processed by competent authorities for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties.

The distinction is particularly relevant within collaborative research and innovation projects, where identical datasets or technological components may be processed under different legal regimes at different stages of the AI lifecycle. During system development, model training or technical validation, processing activities may fall within the GDPR, provided the relevant conditions for lawful processing are satisfied. Once substantially similar technologies are deployed by competent authorities for operational investigative purposes, the applicable legal framework may instead become the LED. Consequently, legal qualification should always be conducted with reference to the **specific processing activity** under examination rather than to the technological solution as a whole.

Legal qualification further requires identifying the respective responsibilities of the actors involved. Under data protection law, accountability depends upon determining which entity acts as **controller**, joint controller or processor, according to the degree of influence exercised over the purposes and essential means of processing.[27] Under the AI Act, accountability is distributed among additional categories of actors, including providers, deployers, authorised representatives, importers, and distributors, each assuming different obligations according to their role within the AI value chain. These categories should not be conflated. A technology provider may simultaneously act as a controller for certain processing activities under the GDPR while assuming the role of provider under the AI Act, whereas a LEA deploying the same system may become both controller under the LED and deployer under the AI Act. Correct governance therefore requires analysing these legal roles independently, without assuming a one-to-one correspondence between regulatory frameworks.[28]

Once responsibilities have been identified, the assessment proceeds to the legal conditions governing processing. Particular attention should be devoted to the existence of an appropriate legal basis, the competence of the authority performing the processing, compliance with the principles of purpose limitation, data minimisation, necessity and proportionality, as well as the establishment of appropriate retention periods and lawful arrangements for data sharing where multiple organisations participate in investigative activities. These principles represent fundamental safeguards ensuring that AI-enabled analytical capabilities remain aligned with the rule of law irrespective of their technological sophistication. In practice, the availability of advanced analytical functions cannot compensate for the absence of a valid legal basis or justify the use of data beyond the purposes for which they were originally collected.

[27] European Data Protection Board, "*Guidelines 07/2020 on the concepts of controller and processor in the GDPR*", 2020. Guidelines 07/2020 on the concepts of controller and processor in the GDPR | European Data Protection Board.
[28] GDPR, Articles 4(7), 4(8), 26 and 28; AI Act, Articles 3 and 25–29 concerning the allocation of responsibilities among providers, deployers and other operators.

Within this governance framework, the AI Act introduces an additional layer of analysis that complements the existing legal obligations. Qualification as an AI system, and where relevant as a high-risk AI system, should never be inferred solely from the fact that a technology is deployed by a law enforcement authority. Instead, the assessment should be conducted through a structured examination of the system's intended purpose, core functionality, the nature of its outputs, the extent to which those outputs influence downstream decision-making and the potential impact on natural persons. This approach is consistent with the risk-based architecture of the AI Act and avoids the over-inclusive assumption that every AI-enabled tool used within law enforcement automatically falls within the scope of high-risk regulation.[29]

Equally important is recognising that legal qualification is **not a one-off exercise**. AI-enabled intelligence-support systems frequently evolve through software updates, integration with additional datasets, changes in operational workflows or the gradual expansion of their use beyond the context originally envisaged by developers. Such developments may alter the applicable legal framework, modify the allocation of responsibilities or introduce new obligations concerning risk management, human oversight, documentation or post-deployment monitoring. Governance should therefore incorporate periodic reassessment of the legal qualification throughout the system lifecycle, ensuring that regulatory compliance evolves alongside technological and operational change.

Ultimately, legal qualification represents the foundation upon which every subsequent governance activity examined in this report is built. Determining the applicable legal regime, identifying accountable actors and clearly defining the intended purpose of the technology are not merely preliminary compliance tasks; they establish the conditions under which impact assessments, safeguards, oversight mechanisms and accountability arrangements can operate effectively. Without this initial qualification, even technically sophisticated AI systems risk being deployed within governance frameworks that are legally uncertain, organisationally fragmented and insufficiently capable of protecting the fundamental rights upon which public trust ultimately depends.

FUNDAMENTAL RIGHTS, IMPACT ASSESSMENT AND VULNERABLE GROUPS

The governance of AI-enabled intelligence-support technologies requires a multidimensional assessment extending beyond legal compliance alone. As discussed in the previous sections, the deployment of AI systems within the FCT domain may affect individuals, groups and society through multiple pathways that cannot be adequately captured by a single assessment instrument. Data protection, fundamental rights, ethical governance, and societal legitimacy therefore represent complementary dimensions of the same governance process, each addressing different aspects of the risks associated with AI-supported investigative activities. From this perspective, the various impact assessment instruments should be regarded as an integrated rights-based governance model, each addressing a specific dimension of AI deployment. Their respective objectives are summarised in Table 1.

[29] AI Act, Article 6 and Annex III; European Commission, *Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689.*
European Commission, *"Draft Commission guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689 (AI Act) for stakeholder consultation"*. Draft Commission guidelines on the classification of high-risk AI systems | Shaping Europe's digital future.
See also Europol, *Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making*, which emphasises that governance should be determined by the operational context and intended use of the technology rather than by the technology alone.

Table 1: Complementary impact assessment instruments within AI-enabled intelligence governance

ASSESSMENT INSTRUMENT	PRIMARY FOCUS	TYPICAL QUESTIONS ADDRESSED
Data Protection Impact Assessment (GDPR, Art. 35)	Risks to the rights and freedoms of natural persons arising from personal data processing	Is the processing lawful, necessary and proportionate? Which technical and organisational measures mitigate privacy and data protection risks?
Law Enforcement Impact Assessment (LED, Art. 27)	Risks associated with processing by competent authorities for criminal law enforcement purposes	Does the processing remain necessary and proportionate for law enforcement purposes? Are appropriate safeguards implemented throughout the processing lifecycle?
Fundamental Rights Impact Assessment (FRIA, AI Act, Art. 27)	Broader impacts on fundamental rights beyond data protection	Which rights may be affected? Who may be affected? How significant are the impacts? Are the safeguards sufficient and proportionate?
Ethical and Societal Impact Assessment	Impacts on public trust, legitimacy, fairness, accountability and societal values	Is the deployment ethically justified? Could societal harms arise despite legal compliance? Are vulnerable groups disproportionately affected?

Within the European legal framework, impact assessment constitutes one of the principal mechanisms through which organisations demonstrate accountability before, during and throughout the deployment of high-impact technologies. However, the purpose of these assessments differs according to the applicable legal regime and the interests being protected. Data Protection Impact Assessments under Article 35 of the GDPR focus primarily on risks to the rights and freedoms of natural persons arising from the processing of personal data, while Article 27 of the LED requires competent authorities to assess the implications of processing operations that are likely to result in high risks within criminal law enforcement contexts. Neither instrument, however, is intended to exhaustively address the broader societal, ethical and constitutional implications that AI-enabled technologies may generate.

For this reason, recent European regulatory developments increasingly advocate complementing traditional privacy-oriented assessments with broader Fundamental Rights Impact Assessments (FRIAs) and structured ethical evaluation methodologies. These approaches examine how AI deployment may affect the effective enjoyment of fundamental rights within specific operational contexts. In particular, the emerging FRIA methodology proposed in both academic literature and European AI governance initiatives emphasises that impact assessment should not be limited to identifying which rights may theoretically be affected. Instead, it should systematically analyse the operational context, the categories of individuals concerned, the severity and likelihood of potential impacts, the safeguards implemented, the residual risks remaining after mitigation and the overall justification for deploying the technology.[30]

[30] Mantelero, A., The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template, 2024. *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template - ScienceDirect*.
ALIGNER Project, Fundamental Rights Impact Assessment. *Fundamental Rights Impact Assessment (FRIA) | aligner*. Regulation (EU) 2024/1689 (AI Act), including provisions concerning Fundamental Rights Impact Assessments where applicable and the Draft Commission Guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689, which reinforce the contextual assessment of intended purpose, downstream use and foreseeable impacts on natural persons.

This contextual perspective is particularly relevant for intelligence-support technologies, whose impacts frequently extend beyond the direct processing of personal data. AI-supported analytical workflows may influence investigative priorities, shape resource allocation, affect evidential interpretation or alter the degree of scrutiny applied to particular individuals or communities. Consequently, the assessment should not focus exclusively on the technology itself but on the broader socio-technical environment within which AI-generated outputs are produced, interpreted and relied upon by human decision-makers.

Several fundamental rights require particular attention within this context. **Privacy and data protection**[31] remain central because AI-enabled analytical systems frequently aggregate information originating from multiple datasets, increasing the risk of function creep, excessive retention, re-identification or secondary uses that exceed the purposes for which information was originally collected. Equally important are **freedom of expression and freedom of information**, particularly where large-scale monitoring of publicly accessible online environments may discourage lawful participation in democratic debate or create chilling effects among journalists, researchers, civil society organisations or human rights defenders.[32] These risks arise not necessarily because information is collected unlawfully, but because systematic aggregation and AI-supported analysis may substantially alter the practical consequences of otherwise legitimate public activities.

The assessment should also carefully consider **equality and non-discrimination**. AI systems may reproduce or amplify existing structural inequalities where historical datasets contain embedded biases, where proxies indirectly correlate with protected characteristics or where analytical outputs disproportionately affect specific communities despite apparently neutral technical design. As highlighted in the academic literature, statistical notions of algorithmic fairness do not automatically correspond to the legal standards governing discrimination under European law. Consequently, fairness assessment requires legal, contextual and organisational analysis in addition to technical performance evaluation.[33]

Similarly, AI-enabled intelligence-support technologies may influence procedural guarantees, including the presumption of innocence, due process and the right to an effective remedy. Analytical outputs such as risk scores, anomaly alerts, relationship graphs or prioritisation mechanisms are generally intended to support rather than replace human decision-making. Nevertheless, where these outputs are afforded excessive authority or insufficiently scrutinised, they may indirectly influence investigative actions affecting individuals who are subsequently unable to understand, challenge or effectively contest the reasoning underpinning AI-supported analytical conclusions. Ensuring meaningful human oversight therefore contributes not only to technical reliability, but also to safeguarding procedural fairness and democratic accountability.

[31] Charter of Fundamental Rights of the European Union, Articles 7 and 8; European Convention on Human Rights, Article 8; Regulation (EU) 2016/679 (GDPR); Directive (EU) 2016/680 (Law Enforcement Directive); Council of Europe Convention 108+, where applicable.

[32] Charter of Fundamental Rights of the European Union, Article 11; European Convention on Human Rights, Article 10. See also European Court of Human Rights, Big Brother Watch and Others v. the United Kingdom [GC], Applications Nos. 58170/13, 62322/14 and 24960/15, Judgment of 25 May 2021, recognising the implications of large-scale surveillance measures for freedom of expression and journalistic activities.

[33] Wachter, S., Mittelstadt, B., & Russell, C., Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-discrimination Law and AI.

The assessment should finally consider broader questions relating to **human dignity and collective or societal harms**. Certain impacts cannot be fully understood through an individual-rights perspective alone. Large-scale analytical systems may contribute to changes in social behaviour, reduced public confidence in law enforcement institutions, unequal distribution of surveillance burdens or the gradual normalisation of intrusive investigative practices. These collective effects are particularly significant because they may materialise even where no individual legal violation can be readily identified. Ethical and societal impact assessment therefore complements traditional legal analysis by examining the cumulative consequences of technological deployment for democratic institutions, public legitimacy and social trust.

Particular attention should also be devoted to **vulnerable groups**, recognising that AI-enabled investigative technologies do not affect all individuals equally. Children and minors may require enhanced protection because of their limited capacity to understand or challenge the consequences of automated analytical processes. Victims and witnesses may experience secondary harms where sensitive personal information is unnecessarily retained, combined or analysed beyond the immediate needs of an investigation. Journalists, activists and members of civil society organisations may be disproportionately affected by monitoring activities capable of interfering with freedom of expression, freedom of association or the confidentiality of professional activities. Members of minority communities may face indirect discrimination where analytical models reflect historical policing patterns or other structural biases. Finally, individuals incidentally captured during intelligence collection – despite having no connection with criminal activities – may nevertheless experience adverse consequences through data aggregation, relationship analysis or erroneous analytical associations.

All the above considerations demonstrate that effective impact assessment cannot be reduced to a checklist of legal obligations or a catalogue of fundamental rights. On the contrary, it should function as a structured governance process capable of linking the operational context, the individuals affected, the rights and interests at stake, the probability and severity of potential impacts, the safeguards implemented, the residual risks remaining after mitigation, and the reasons justifying the deployment of the technology. This integrated approach reflects the growing convergence between European data protection law, the AI Act, ethical AI governance and the evolving doctrine on FRIA. Ultimately, the objective is not merely to demonstrate regulatory compliance, but to ensure that AI-enabled intelligence-support technologies remain lawful, proportionate, ethically justified and socially legitimate throughout their operational lifecycle.

All the above considerations demonstrate that effective impact assessment cannot be reduced to a checklist of legal obligations or a catalogue of fundamental rights. On the contrary, it should function as a structured governance process capable of linking the operational context, the individuals affected, the rights and interests at stake, the probability and severity of potential impacts, the safeguards implemented, the residual risks remaining after mitigation, and the reasons justifying the deployment of the technology. This integrated approach reflects the growing convergence between European data protection law, the AI Act, ethical AI governance and the evolving doctrine on FRIA. Ultimately, the objective is not merely to demonstrate regulatory compliance, but to ensure that AI-enabled intelligence-support technologies remain lawful, proportionate, ethically justified and socially legitimate throughout their operational lifecycle.

ETHICAL DECISION-MAKING AND PUBLIC LEGITIMACY

The legal and governance assessments discussed in the previous sections provide the foundation for the responsible deployment of AI-enabled intelligence-support technologies. Nevertheless, compliance with applicable legislation and the implementation of appropriate safeguards do not necessarily resolve every decision that organisations may face in practice. Situations frequently arise in which several lawful options remain available, each involving different implications for individuals, operational effectiveness and public trust. Ethical governance therefore complements legal compliance by providing a structured process through which organisations can justify, document and communicate decisions whose consequences extend beyond purely legal considerations.

Recognising this challenge, Europol has developed a practical methodology for ethical decision-making specifically designed for technology assessment within the law enforcement domain. The methodology structures ethical reasoning around a sequence of documented questions that connect operational realities with the values underpinning democratic policing. This approach acknowledges that legality and ethical acceptability, although closely related, do not necessarily coincide. A technologically lawful solution may still generate disproportionate societal impacts or undermine public confidence if broader ethical considerations are not adequately addressed.[34]

The Europol methodology can be operationalised through the decision-making framework summarised in Table 2.

Table 2: Ethical decision-making framework adapted from the Europol methodology

DECISION STEP	KEY GOVERNANCE QUESTION
Moral problem	What is the ethical issue arising from the proposed deployment or operational use of the AI system?
Relevant facts	What technical, legal, operational and organisational information is necessary to understand the situation?
Affected parties	Which individuals, groups or institutions may benefit from or be adversely affected by the decision?
Values at stake	Which ethical values and fundamental rights are potentially in tension?
Available options	What realistic alternatives exist, including the option not to deploy or to limit deployment?
Justification	Why is the preferred option ethically and operationally justified, taking account of safeguards and proportionality?
Consequences and review	How will the decision be documented, monitored and reviewed as circumstances evolve?

[34] Europol, "*Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making*"; High-Level Expert Group on *Artificial Intelligence, Ethics Guidelines for Trustworthy AI* (ALTAI).

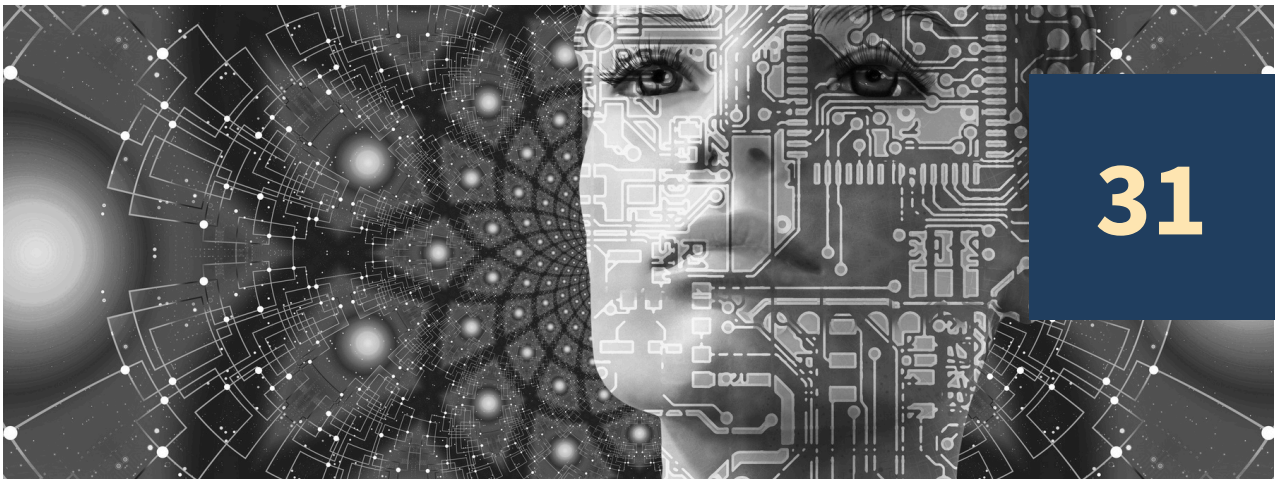
Applying this framework encourages decision-makers to move beyond a purely technical assessment of system performance and to consider the broader implications of deployment throughout the operational lifecycle. Ethical reasoning should therefore integrate the principal values recognised by European AI governance frameworks, including transparency, fairness, privacy, accountability, human dignity, autonomy, justice, and public legitimacy. These values should not be regarded as abstract aspirations but as practical criteria guiding organisational choices whenever competing interests cannot be resolved solely through legal interpretation.

In operational environments, ethical dilemmas frequently arise where these values must be balanced against one another. Expanding data collection may improve investigative capabilities while simultaneously increasing risks to privacy. Highly sensitive analytical models may enhance the detection of criminal activities but also increase false positives affecting innocent individuals. Automated prioritisation may improve operational efficiency yet reduce transparency regarding the rationale underpinning investigative decisions. Structured ethical deliberation helps ensure that such trade-offs are explicitly identified, justified, and documented rather than being addressed implicitly or left to individual discretion.

The consideration of affected parties is particularly important within AI-supported intelligence activities. As highlighted in the previous section, the consequences of deployment often extend beyond suspects or persons directly involved in criminal investigations. Victims, witnesses, children, journalists, civil society organisations, minority communities, operational personnel and individuals incidentally captured during intelligence collection may all experience different forms of impact. Ethical assessment therefore complements legal and fundamental-rights analysis by examining how competing interests are distributed across these different stakeholders and whether less intrusive alternatives could achieve comparable operational objectives.

Ultimately, ethical decision-making contributes directly to **public legitimacy**, which represents one of the essential conditions for the sustainable use of AI within democratic societies. Public confidence in law enforcement does not depend solely upon the legality of technological tools, but also upon the perception that their deployment is fair, transparent, proportionate and subject to meaningful accountability. Documenting ethical reasoning, together with the legal and fundamental-rights assessments described throughout this section, strengthens organisational accountability by demonstrating not only that a particular course of action was legally permissible, but also that it was consciously evaluated against the broader values that underpin the rule of law. In this sense, ethical governance functions as the bridge between regulatory compliance and trustworthy AI deployment, reinforcing the legitimacy of operational decisions throughout the lifecycle of AI-enabled intelligence-support technologies.

[34] Europol, “Assessing Technology in Law Enforcement: A Method for Ethical Decision-Making”; High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (ALTAI).



5 - IMPACT ON FCT R&I LANDSCAPE: OPERATIONAL SAFEGUARDS AND EFFECTIVENESS METRICS

THE FOUR-LAYER SAFEGUARDS MODEL

The analyses presented in the previous sections have shown that the governance of AI-enabled intelligence-support technologies cannot rely on isolated safeguards addressing individual legal, technical or ethical issues. Risks emerge throughout the entire AI-supported investigative lifecycle and often propagate across multiple organisational and operational levels. Consequently, effective governance requires an integrated safeguards architecture capable of linking legal compliance, fundamental rights protection, operational reliability, and public accountability within a single governance framework.

Building upon the Rights-Based Operational Governance Methodology introduced in the Analytical Methodology and the legal and ethical analyses developed in section 4, this report proposes a four-layer safeguards model. The model organises safeguards into complementary governance layers that collectively support the lawful, trustworthy and accountable deployment of AI-enabled intelligence-support technologies. While legal and ethical considerations should be addressed throughout the AI lifecycle, the emphasis on deployment reflects the fact that many governance, accountability and fundamental rights challenges materialise when AI systems are embedded in operational environments, and their outputs influence investigative activities and decision-making.

This distinction is particularly relevant in the FCT research and innovation domain, where technologies developed and validated during research projects may subsequently be deployed by Law Enforcement Authorities under different legal, organisational and operational conditions. Each layer addresses a distinct governance objective while reinforcing the effectiveness of the others, creating a continuous governance cycle.

The **first layer** establishes the **legal and organisational foundations** upon which every subsequent safeguard depends. Before any analytical capability is deployed, organisations should demonstrate the existence of an appropriate legal authority, clearly define the intended operational purpose of the system, and explicitly identify any excluded uses that fall outside the authorised scope of deployment. This layer further requires documenting the necessity and proportionality of the envisaged processing, defining retention periods consistent with the applicable legal framework, restricting access according to operational responsibilities and establishing appropriate conditions governing data sharing between participating organisations. These safeguards ensure that technological capabilities remain aligned with clearly defined legal objectives throughout the system lifecycle and reduce the risk of function creep or unauthorised secondary uses.

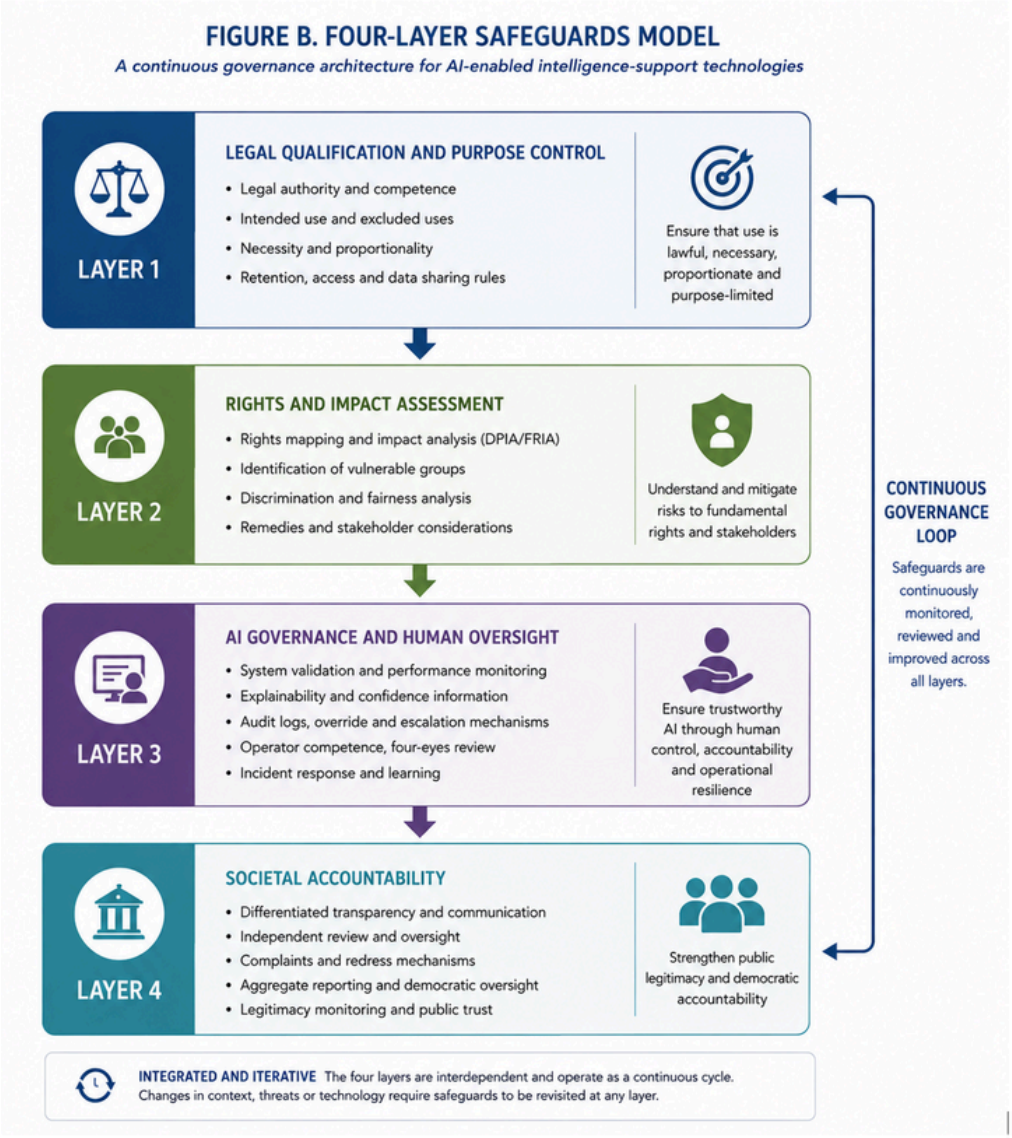


*Figure 2: Four-layer safeguards model for AI-enabled intelligence-support technologies**

The **second layer** focuses on understanding **how AI-supported activities may affect individuals, groups and society**. As discussed in Section 4, governance should extend beyond data protection considerations alone and incorporate a broader assessment of fundamental rights and societal impacts. Safeguards at this level therefore include systematic rights mapping, the identification of potentially affected stakeholders and vulnerable groups, the completion of the appropriate impact assessments (such as DPIAs, Article 27 LED assessments or FRIAs, where applicable), and structured analyses of discrimination risks and available remedies. These assessments provide evidence supporting the continuous evaluation of whether deployment remains justified as operational conditions evolve.

The **third layer** translates legal and ethical requirements into **day-to-day operational practices** governing the interaction between AI systems and human operators. At this level, safeguards concentrate on ensuring that AI-generated outputs remain technically reliable, understandable and subject to meaningful human control. Organisations should therefore implement documented validation procedures, provide appropriate confidence information and explanatory features where technically feasible, maintain comprehensive audit logs and establish mechanisms allowing operators to challenge, override or escalate AI-supported recommendations whenever necessary.

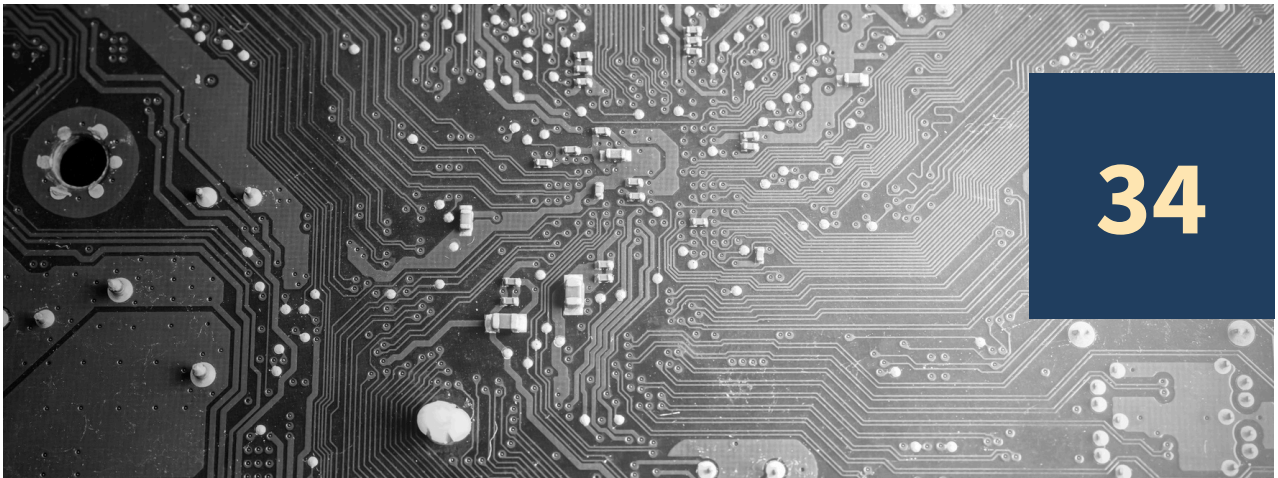
* Author's own elaboration, informed by the risk categories and governance framework discussed in this report.

Effective governance further depends upon ensuring that personnel possess the competence required to interpret analytical outputs critically, that significant decisions benefit from appropriate peer review (such as the application of a four-eyes principle where operationally justified) and that incidents, anomalies and unexpected system behaviour are systematically recorded, investigated and used to improve future system performance.

The **fourth layer** extends governance beyond the immediate operational environment and addresses the **broader accountability obligations** that AI-enabled intelligence-support technologies generate **within democratic societies**. Safeguards at this level include differentiated transparency measures adapted to the needs of different stakeholders, mechanisms for independent oversight and review, accessible complaint procedures, periodic aggregate reporting on system use and performance, appropriate communication with the public and continuous monitoring of public legitimacy. These measures recognise that trustworthy AI cannot be achieved solely through internal organisational controls but also depends upon maintaining external accountability and preserving confidence in the institutions responsible for deploying advanced analytical technologies.

All four layers should not be interpreted as successive stages completed once and then abandoned. Instead, they form a dynamic governance architecture in which legal qualification, impact assessment, operational oversight, and societal accountability continuously inform one another throughout the AI lifecycle. Changes in operational objectives, datasets, technological capabilities, threat environments or regulatory requirements may require safeguards to be revisited at any layer. Consequently, governance should be understood as an iterative process supported by continuous monitoring, periodic review and evidence-based improvement rather than as the achievement of a fixed state of compliance.

An important characteristic of the proposed model is the **distinction between implemented safeguards, planned improvements, and residual risks**. Governance should not assume that the existence of policies or future commitments demonstrates effective risk mitigation. Instead, organisations should be able to distinguish safeguards that are already evidenced through documentation, technical implementation or operational practice from additional measures that remain under development. Likewise, the effectiveness of implemented safeguards should be evaluated against the residual risks that continue to exist despite mitigation efforts. Maintaining this distinction enhances organisational transparency, supports accountability and enables governance decisions to be based on demonstrable evidence rather than assumptions regarding future implementation.



MEANINGFUL HUMAN OVERSIGHT

Meaningful human oversight represents one of the central governance mechanisms for ensuring that AI-enabled intelligence support systems remain compatible with fundamental rights, professional accountability, and the rule of law. However, both the EDPS and the broader interdisciplinary literature[35] consistently emphasise that human oversight should not be regarded as a universal remedy capable of legitimising otherwise problematic systems. Simply placing a human operator somewhere within the decision-making chain does not automatically prevent unlawful processing, discriminatory outcomes or excessive reliance on automated outputs. Rather, the effectiveness of human oversight depends on the way technical design, organisational processes, and institutional culture interact throughout the operational lifecycle.[36]

For intelligence-support technologies used in the FCT domain, meaningful human oversight should therefore be understood as a multidimensional governance capability and not as a single procedural requirement. It requires that operators possess both the authority and the practical ability to critically evaluate AI-generated outputs, question automated recommendations where appropriate, and intervene whenever risks to legality, proportionality or investigative reliability emerge. This perspective is consistent with the AI Act, which frames human oversight as an integral characteristic of high-risk AI systems instead of an optional safeguard added after deployment.[37]

A first dimension concerns **governance arrangements**. Human oversight can only be meaningful when operators have clearly defined responsibilities, appropriate competence and genuine independence in exercising professional judgement. Personnel responsible for validating AI-supported analytical outputs should receive role-specific training covering not only system functionality, but also known limitations, sources of uncertainty, potential biases, and applicable legal obligations. Equally important, organisational structures should ensure that analysts are not subject to inappropriate operational pressure to follow algorithmic recommendations solely because they originate from automated systems. Accountability therefore requires clearly allocated responsibilities, documented decision-making authority, and explicit recognition that responsibility for operational decisions always remains with the competent authority, rather than with the technology itself.[38]

The second dimension relates to **operational conditions** under which oversight takes place. Even highly qualified personnel cannot exercise effective judgement if operational workflows provide insufficient time to review system outputs or create excessive workload that encourages automatic acceptance of algorithmic recommendations. Meaningful oversight therefore depends on realistic staffing levels, proportionate workloads, and procedures that allow operators to examine the evidential basis supporting AI-generated alerts before these influence investigative priorities. Depending on the operational context and associated risks, additional organisational safeguards (such as peer validation, supervisory review or application of a four-eyes principle) may further strengthen the reliability of decisions affecting individuals' rights. Clear escalation procedures should also be established whenever uncertainty exceeds predefined thresholds, system behaviour appears anomalous or significant legal or ethical concerns arise.

[35] See European Data Protection Supervisor (EDPS), Checklist on Human Intervention in AI-supported Decision-Making; EDPS, TechDispatch: Human Oversight (2023); Sterz, A., et al. (2024), On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives, Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency; High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (2019); Regulation (EU) 2024/1689 (AI Act), in particular Article 14 and Recitals 27, 73 and 92.

[36] European Data Protection Supervisor (EDPS), "Guidelines on generative AI: embracing opportunities, protecting people", 2024. EDPS-2024-09-Generative-AI-guidelines_EN.pdf.

[37] Regulation (EU) 2024/1689 (AI Act), provisions on human oversight and obligations of deployers/users of high-risk AI systems (see articles: 13, 14, 15 and 29).

[38] NIST, AI Risk Management Framework. AI Risk Management Framework | NIST.

A third dimension concerns the **technical environment** supporting human decision-making. Interface design should facilitate critical assessment, avoiding passive acceptance of automated recommendations. Operators should have access to information that enables them to understand the basis of AI-supported outputs, including confidence levels, data provenance, known system limitations and any indicators of uncertainty relevant to the analytical result. Logging mechanisms should record both system-generated outputs and subsequent human interventions, creating an auditable record of how recommendations were assessed, accepted, modified or rejected. In particular, explainability mechanisms should provide information that is meaningful for operational decision-making and proportionate to the context in which the system is deployed. These technical measures reduce the likelihood that users place unwarranted confidence in automated outputs while simultaneously strengthening transparency and accountability.[39]

Particular attention should be given to **automation bias**, namely the well-documented tendency of individuals to overestimate the reliability of automated recommendations or to disregard contradictory evidence once an algorithm has produced a seemingly authoritative result. Empirical studies have shown that automation bias may occur even among experienced professionals and can be reinforced by organisational cultures that prioritise efficiency over critical reflection.[40] In intelligence environments characterised by time pressure, high information volumes, and cognitive overload, operators may increasingly rely on algorithmic outputs as cognitive shortcuts, thereby reducing independent analytical judgement. Effective governance therefore requires continuous awareness training, interface designs that promote active verification and organisational cultures that explicitly encourage questioning automated recommendations whenever uncertainty or inconsistency arises.[41]

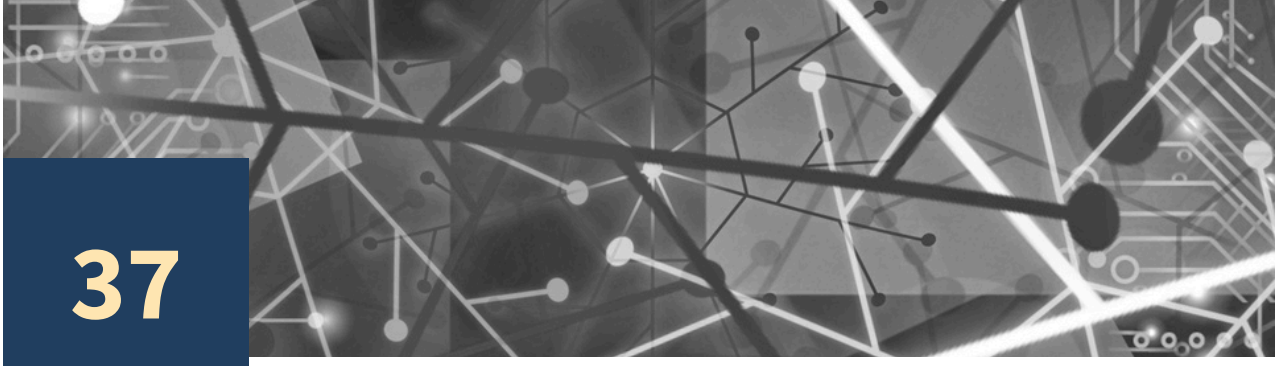
The final dimension concerns **continuous organisational learning**. Human oversight does not end once an operator validates or rejects an AI-generated output. Overridden decisions should be systematically documented and periodically reviewed to identify recurring patterns, false positives, usability issues or emerging sources of error. Root-cause analysis following significant incidents can reveal whether shortcomings originate from training data, model behaviour, interface design, operational procedures or organisational practices. Lessons learned should subsequently feed into model retraining where appropriate, procedural updates, operator training programmes, and broader governance reviews. In this way, human oversight becomes a dynamic feedback mechanism capable of improving both technical performance and organisational resilience over time, rather than merely acting as a final approval stage before operational use. This continuous improvement approach closely reflects contemporary AI governance frameworks, which increasingly view effective oversight as an iterative process embedded throughout the entire AI lifecycle and not as a single point of human intervention.[42]

[39] Spanish Data Protection Agency (AEPD), Audit Requirements for Personal Data Processing Activities Involving AI, 2021. Audit Requirements for Personal Data Processing Activities involving AI.

[40] See Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 39(2), 230–253; Mosier, K. L., Skitka, L. J., Burdick, M., & Heers, S. (1996). Automation Bias: Decision Making and Performance in High-Tech Cockpits. International Journal of Aviation Psychology, 6(1), 47–63; Skitka, L. J., Mosier, K. L., & Burdick, M. D. (2000). Does Automation Bias Decision-Making? International Journal of Human-Computer Studies, 52(5), 991–1006; Cummings, M. L. (2004). Automation Bias in Intelligent Time-Critical Decision Support Systems. AIAA First Intelligent Systems Technical Conference. See also European Data Protection Supervisor (EDPS), Checklist on Human Intervention in AI-supported Decision-Making, and Sterz, A. et al. (2024), On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives, Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency.

[41] Information Commissioner's Office, "How do we ensure fairness in AI?". How do we ensure fairness in AI? | ICO.

[42] NIST, "TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management", 2022. TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management - December 1, 20222.



In conclusion, human oversight effectiveness depends on the alignment of governance structures, operational procedures, technical design and continuous organisational learning. Only when these elements operate coherently can human oversight fulfil its intended function of preserving human agency, ensuring accountability and maintaining public trust in the use of AI-enabled intelligence support technologies within the FCT domain.

SAFEGUARDS EFFECTIVENESS MATRIX

The governance measures discussed throughout this report demonstrate that the existence of safeguards alone does not necessarily ensure that AI-enabled intelligence-support technologies operate lawfully, fairly or effectively. Policies may exist without being implemented, technical controls may be activated without being regularly verified, and human oversight may formally satisfy regulatory requirements while failing to exercise meaningful influence over operational decisions. Consequently, the assessment of AI governance should extend beyond the mere presence of safeguards and examine whether these safeguards demonstrably reduce the risks they are intended to address.

To support this objective, this report proposes a **Safeguards Effectiveness Matrix** as a practical governance instrument for monitoring, reviewing and continuously improving AI-enabled intelligence-support systems throughout their operational lifecycle. The matrix provides a structured mechanism for linking each safeguard to the specific risks it addresses, the actors responsible for its implementation, the evidence demonstrating its operation, and the indicators that can be used to evaluate its continuing effectiveness.

This evidence-based perspective is consistent with the accountability principle underpinning both the GDPR and the LED, as well as with the lifecycle approach adopted by the AI Act and contemporary AI governance frameworks. Effective governance therefore requires organisations not only to implement safeguards, but also to demonstrate, through documented evidence, that these safeguards remain proportionate, operationally effective, and responsive to evolving risks. Regular monitoring further enables organisations to identify declining performance, emerging vulnerabilities or unintended consequences before they result in operational failures or infringements of fundamental rights.[43]

An important characteristic of the proposed matrix is the integration of both **quantitative and qualitative indicators**. While technical performance can often be assessed through measurable metrics (such as false-positive rates, override frequency or subgroup error rates), many governance objectives cannot be adequately evaluated through numerical indicators alone. The assessment of legal proportionality, the adequacy of remedies, the sufficiency of human oversight or the legitimacy of operational decisions inevitably requires legal interpretation and contextual judgement. In fact, statistical measures of fairness cannot replace legal reasoning concerning discrimination, equality or procedural justice. Technical metrics therefore provide evidence supporting governance decisions, but they do not determine legal compliance or ethical acceptability in isolation.

[43] Regulation (EU) 2016/679 (GDPR), Articles 5(2), 24 and 35; Directive (EU) 2016/680, Articles 4, 19 and 27; Regulation (EU) 2024/1689 (AI Act), Articles 9, 14, 15, 17 and 29.

The matrix also introduces explicit **governance triggers**. Monitoring indicators have limited value unless organisations establish predefined conditions requiring review or escalation. Thresholds may be quantitative – for example, an increase in false-positive rates beyond an acceptable level – or qualitative, such as repeated operator concerns regarding unexplained outputs, evidence of function creep or significant changes in operational context. Once predefined thresholds are exceeded, governance mechanisms should require additional validation, legal reassessment, model retraining, procedural revision or, where necessary, temporary suspension of operational deployment until identified risks have been adequately addressed.

Finally, the matrix incorporates a **maturity assessment** for each safeguard, recognising different stages of implementation. A safeguard may be absent, formally defined but not yet operational, effectively implemented, continuously monitored or fully validated through periodic review and evidence-based evaluation. This maturity perspective supports continuous improvement by enabling organisations to identify governance gaps, prioritise resource allocation, and demonstrate progressive enhancement of organisational capabilities over time.

Table 3 presents a representative Safeguards Effectiveness Matrix that can be adapted to different AI-enabled intelligence-support applications while preserving a consistent governance structure across operational contexts. For each safeguard, a single maturity level should be assigned on the basis of the available evidence, reflecting its current state of implementation rather than its intended future status.

Table 3: Safeguards Effectiveness Matrix for AI-enabled intelligence-support technologies

SAFEGUARD	RISK ADDRESSED	RESPONSIBLE ACTOR	EVIDENCE	INDICATOR	THRESHOLD / TRIGGER	REVIEW CYCLE	CURRENT MATURITY LEVEL
Documented intended purpose	Function creep; unlawful processing	Controller/ Deployer	Governance documentation	% of use cases with documented intended purpose	New operational purpose identified	Annual / Event-driven	Absent (A) Defined (D) Implemented (I) Monitored (M) Validated (V)
Documented legal basis	Unlawful processing	Controller	Legal assessment	% of deployments with documented legal basis	Legal uncertainty or legislative change	Annual	A–D–I–M–V
Data provenance verification	Poor data quality; unreliable outputs	Provider / Controller	Data provenance records	% of outputs supported by provenance evidence	Missing or unverifiable sources	Continuous	A–D–I–M–V
Confidence information available	Automation bias	Provider	System documentation	Visibility of confidence information	Confidence unavailable or inconsistent	Continuous	A–D–I–M–V
Override mechanism	Excessive automation	Deployer / Operator	Audit logs	Override frequency	Repeated inability to override	Continuous	A–D–I–M–V
Operator competence and training	Misinterpretation of outputs	Deployer	Training records	Training completion rate	Training expired or incomplete	Annual	A–D–I–M–V
Escalation procedure	Unresolved anomalies	Operator / Supervisor	Incident register	Escalation rate	Repeated anomalies or uncertainty	Event-driven	A–D–I–M–V

SAFEGUARD	RISK ADDRESSED	RESPONSIBLE ACTOR	EVIDENCE	INDICATOR	THRESHOLD / TRIGGER	REVIEW CYCLE	CURRENT MATURITY LEVEL
Audit logging	Lack of accountability	Provider / Deployer	System logs	Decision-log completeness	Missing audit trail	Continuous	A–D–I–M–V
Bias monitoring	Discriminatory outcomes	Provider / Controller	Validation reports	Subgroup error rate	Significant performance disparity	Periodic	A–D–I–M–V
Incident reporting and response	Operational failures	Deployer	Incident reports	Incident closure time	Serious incident detected	Event-driven	A–D–I–M–V
Periodic legal and ethical review	Regulatory drift	Controller / DPO / Ethics Board	Review reports	Completion of scheduled reviews	Legislative or operational change	Annual	A–D–I–M–V
Transparency and accountability reporting	Loss of public trust	Controller	Public reports	Publication of governance report	Failure to publish scheduled report	Annual	A–D–I–M–V
Complaints and redress mechanism	Lack of effective remedy	Controller / Oversight body	Complaint register	Resolution time and adequacy of remedy	Repeated unresolved complaints	Periodic	A–D–I–M–V
Bias monitoring	Discriminatory outcomes	Provider / Controller	Validation reports	Subgroup error rate	Significant performance disparity	Periodic	A–D–I–M–V
Incident reporting and response	Operational failures	Deployer	Incident reports	Incident closure time	Serious incident detected	Event-driven	A–D–I–M–V
Periodic legal and ethical review	Regulatory drift	Controller / DPO / Ethics Board	Review reports	Completion of scheduled reviews	Legislative or operational change	Annual	A–D–I–M–V
Transparency and accountability reporting	Loss of public trust	Controller	Public reports	Publication of governance report	Failure to publish scheduled report	Annual	A–D–I–M–V

6 - OPERATIONAL APPLICATION, VALIDATION AND UPTAKE OF THE GOVERNANCE FRAMEWORK FOR THE FCT R&I ECOSYSTEM

APPLYING THE GOVERNANCE MODEL TO REPRESENTATIVE OPERATIONAL SCENARIOS

The governance model proposed throughout this report is intended to support decision-making across different categories of AI-enabled intelligence-support technologies. To demonstrate its practical applicability, the model is applied below to three representative scenarios that reflect common analytical capabilities currently explored within the FCT R&I ecosystem:

1. **OSINT and social-media intelligence;**
2. **Multilingual and multimodal content analysis; and**
3. **Graph-based intelligence and relationship detection.**

Consistent with the Europol ethical decision-making methodology, these scenarios have been selected not to assess individual technologies, but to illustrate how different operational contexts, affected stakeholders and investigative purposes generate distinct ethical, legal and societal (ELS) challenges despite relying on similar AI capabilities. Accordingly, the assessment starts from the operational context and the persons affected before considering the technological solution itself.

The objective of this application is not to assess the legal compliance of a particular AI system, nor to provide an exhaustive operational risk assessment. Instead, the scenarios illustrate how the safeguards effectiveness model can assist organisations in identifying the relationship between operational risks, affected fundamental rights, governance measures and the evidence required to demonstrate that safeguards remain effective over time. Although each scenario presents different technological characteristics, they also reveal recurring governance challenges, including data quality, transparency, human oversight, accountability, and proportionality.

The scenario documented in Table 4 concerns the collection, filtering and analysis of publicly accessible online information to support intelligence activities and the identification of emerging criminal patterns. Although the underlying information may already be publicly available, its systematic aggregation and AI-assisted analysis may substantially increase its legal and societal significance, requiring careful assessment of proportionality, purpose limitation and accountability.

Table 4: Scenario 1 – OSINT and social-media intelligence

GOVERNANCE ELEMENT	APPLICATION
Primary risk	Collection of inaccurate, outdated or unlawfully obtained information; excessive monitoring; function creep.
Affected (fundamental) rights	Privacy, protection of personal data, freedom of expression, freedom of association, presumption of innocence.
Relevant safeguards	Clearly documented purpose limitation; provenance verification; legal basis assessment; meaningful human oversight; periodic legal review.
Evidence	Source documentation, provenance records, DPIA/LED assessment, audit logs, operator validation records.
Representative indicators	Percentage of outputs supported by verified provenance; override frequency; percentage of monitored sources covered by documented legal basis.
Residual concern	Publicly available information may still generate disproportionate interference when aggregated, analysed or retained over extended periods.

The scenario documented in Table 5 focuses on multilingual and multimodal analytical environments in which text, images, audio and video are processed jointly to support investigations. The integration of heterogeneous information sources increases analytical capabilities but also creates additional challenges related to contextual interpretation, linguistic bias and evidential reliability.

Table 5: Scenario 2 – Multilingual and multimodal content analysis

GOVERNANCE ELEMENT	APPLICATION
Primary risk	Translation errors; cultural bias; incorrect contextual interpretation; unequal performance across languages or modalities.
Affected (fundamental) rights	Equality and non-discrimination; effective defence; fair treatment; freedom of expression.
Relevant safeguards	Multilingual validation; subgroup performance testing; confidence information; operator review; periodic model evaluation.
Evidence	Validation reports; benchmark testing; bias assessments; training documentation; review records.
Representative indicators	Subgroup error rate; false-positive rate across languages; operator disagreement rate; frequency of model retraining.
Residual concern	Contextual nuances, slang, irony or cultural references may remain difficult to capture, requiring continued reliance on human judgement.

The scenario documented in Table 6 concerns graph analytics supporting the identification and visualisation of relationships among persons, organisations and events. AI-generated associations should remain analytical hypotheses requiring independent human verification and should not be interpreted as evidence of criminal responsibility.

Table 6: Scenario 3 – Graph-based intelligence and relationship detection

GOVERNANCE ELEMENT	APPLICATION
Primary risk	False association between individuals; propagation of historical bias; over-reliance on inferred relationships.
Affected (fundamental) rights	Privacy; data protection; non-discrimination; presumption of innocence; right to an effective remedy.
Relevant safeguards	Human validation of inferred links; explainability mechanisms; audit logging; incident reporting.
Evidence	Graph validation records; decision logs; review reports; incident documentation; governance records.
Representative indicators	Decision-log completeness; review completion rate; escalation rate; percentage of inferred links supported by corroborating evidence.
Residual concern	Even technically accurate relationship detection may generate misleading investigative narratives if contextual evidence is incomplete or outdated.

Across the three scenarios, several common governance patterns emerge. AI-enabled intelligence-support technologies affect not only investigators and intelligence analysts, but also a broader ecosystem including suspects, victims, witnesses, minors, journalists, activists, members of minority groups and persons incidentally captured during data collection. Consequently, identical AI functionalities may require different safeguards depending on the operational context, the categories of persons affected and the role assigned to AI outputs within investigative workflows. Consistent with the Europol ethical decision-making methodology, stakeholder mapping therefore remains an essential component of any governance assessment.

First, the most significant risks rarely originate from algorithmic processing alone but instead arise from the interaction between technical limitations, data quality, organisational practices, and human decision-making. **Secondly**, the effectiveness of safeguards depends less on their formal existence than on the availability of objective evidence demonstrating that they are consistently implemented, periodically reviewed and capable of adapting to changing operational conditions. **Finally**, no individual safeguard completely eliminates risks. Governance therefore relies on the cumulative effect of legal, organisational, technical, and accountability measures operating together throughout the AI lifecycle.

The application further illustrates that the proposed model remains sufficiently flexible to accommodate heterogeneous analytical capabilities while maintaining a consistent governance structure. It provides a common framework through which organisations can identify risks, allocate responsibilities, monitor safeguard effectiveness, and support continuous improvement. This adaptability is particularly relevant within the FCT research and innovation ecosystem, where investigative technologies evolve rapidly, and operational contexts differ considerably across organisations and EU Member States.

IMPLICATIONS FOR THE FCT R&I ECOSYSTEM

The governance model developed in this report has implications that extend beyond the assessment of individual AI-enabled intelligence-support systems. Its principal contribution lies in providing a structured methodology that can inform the entire research and innovation lifecycle within the FCT ecosystem, from the initial design of research projects to operational deployment and continuous post-deployment evaluation. By integrating legal analysis, fundamental rights impact assessment, organisational governance, and evidence-based monitoring into a single framework, the model supports a more systematic approach to the responsible development and adoption of AI technologies for law enforcement purposes.

A first implication concerns **research design**. Fundamental rights, data protection, and ethical considerations should not be addressed solely during the validation phase of research projects or immediately before operational deployment. Instead, they should be incorporated from the earliest stages of project conception, influencing research objectives, use-case definition, data selection, system architecture, and evaluation criteria. This approach encourages the adoption of "governance by design," whereby legal, societal and operational requirements evolve alongside technological development. Such an approach also facilitates the identification of foreseeable risks before they become embedded within technical design choices, thereby reducing the need for costly corrective measures during later stages of the project lifecycle.

The proposed methodology also strengthens **project governance** by promoting a multidisciplinary decision-making structure throughout research activities. AI-enabled technologies developed for the FCT domain involve technical, legal, operational and societal considerations that cannot be adequately addressed within disciplinary silos. Effective governance therefore requires continuous interaction among software developers, AI specialists, legal experts, ethicists, operational practitioners, data protection officers, and, where appropriate, representatives of civil society. Therefore, governance becomes a collective process in which technical performance, operational utility and fundamental rights considerations are evaluated together during successive project reviews.

The model is equally relevant to **procurement and innovation uptake**. Public authorities increasingly procure AI-enabled solutions developed by external providers, requiring procurement procedures that extend beyond traditional technical specifications and performance requirements. Governance-related criteria – including transparency of intended purpose, availability of audit documentation, explainability mechanisms, human oversight capabilities, logging functions, and evidence supporting validation activities (as well as, where appropriate, cybersecurity governance and resilience measures consistent with the broader EU cybersecurity framework[44]) – should therefore become integral elements of procurement documentation and supplier evaluation. In this way, procurement evolves from the acquisition of technological products towards the acquisition of demonstrably trustworthy AI-enabled capabilities supported by appropriate governance arrangements.

[44] See Directive (EU) 2022/2555 (NIS2 Directive), in particular Articles 21-23 concerning cybersecurity risk-management measures and supply-chain security. While the NIS2 Directive does not regulate AI governance specifically, it complements the AI Act by strengthening organisational cybersecurity, resilience and governance obligations that are relevant when procuring and deploying AI-enabled systems.

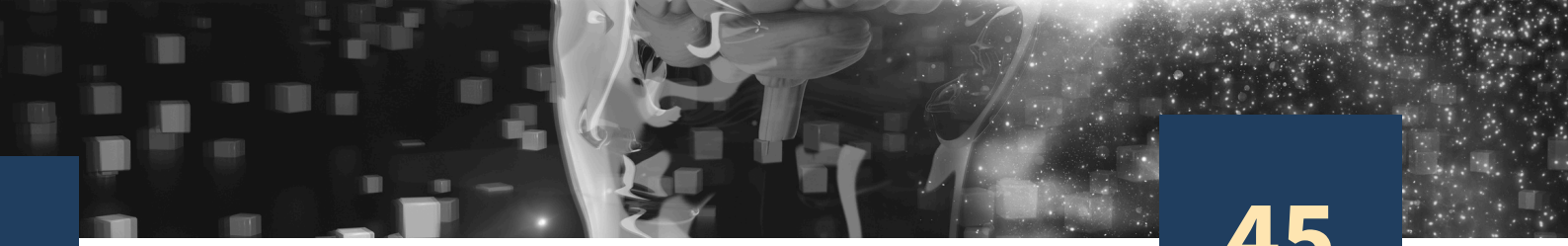
These considerations naturally contribute to **deployment readiness.** Readiness for operational deployment should not be determined solely by technical maturity or functional performance, but also by the organisational capacity to operate AI systems responsibly. Before deployment, organisations should verify that governance arrangements are operational, responsibilities clearly allocated, personnel adequately trained, monitoring procedures established, and evidence collection mechanisms fully functional. The safeguards effectiveness matrix, presented in this report, offers a practical instrument for assessing organisational preparedness by demonstrating not only that safeguards exist, but also that they can be effectively implemented and monitored within the intended operational environment.

The methodology likewise supports ongoing efforts towards **standardisation and interoperability** within the European FCT research ecosystem. One of the recurring challenges across research projects is the absence of common governance criteria that enable consistent comparison between different technological solutions and operational contexts. By introducing common governance components (including safeguards, evidence, indicators, review mechanisms and maturity levels), the model facilitates greater consistency across projects while remaining sufficiently flexible to accommodate diverse technological architectures. Therefore, it provides a common governance vocabulary that can support future standardisation activities, facilitate knowledge transfer between projects and improve comparability across evaluation exercises.

Another critical implication concerns **skills development and organisational competence**. Meaningful governance cannot be achieved solely through technical safeguards or regulatory documentation. It requires personnel capable of understanding both the capabilities and the limitations of AI-enabled intelligence-support technologies. Training programmes should therefore extend beyond system operation and include legal obligations, fundamental rights considerations, algorithmic limitations, automation bias, explainability, evidence evaluation, and incident management. Developing these multidisciplinary competencies contributes to strengthening institutional resilience while supporting more informed operational decision-making.

Finally, the governance framework emphasises the importance of **post-deployment monitoring** as an integral component of responsible AI adoption. Operational environments evolve continuously through changes in investigative priorities, data sources, threat landscapes, and legal requirements. Consequently, governance cannot conclude once deployment has been authorised. Continuous monitoring of safeguard effectiveness, periodic review of legal assumptions, reassessment of operational risks, systematic analysis of incidents, and operator feedback enable organisations to identify emerging risks before they compromise operational effectiveness or fundamental rights. This lifecycle perspective is consistent with contemporary AI governance frameworks, which increasingly recognise that trustworthy AI depends on continuous organisational learning rather than one-off compliance assessments.

Overall, the proposed methodology contributes to strengthening the maturity of the European FCT R&I ecosystem by shifting the focus from isolated compliance activities towards continuous governance. In doing so, it supports the development, procurement and deployment of AI-enabled intelligence-support technologies that are not only technically effective but also demonstrably accountable, transparent and aligned with the rule of law.



VALIDATION AND PRACTICAL UPTAKE

The governance model proposed in this report is intended as a practical decision-support instrument and not as a certification framework or conformity assessment methodology. Its primary objective is to assist organisations in systematically identifying governance requirements, evaluating the effectiveness of safeguards and supporting evidence-based decision-making throughout the lifecycle of AI-enabled intelligence-support technologies. Consequently, validation should focus on assessing the **utility, completeness, usability, and operational relevance** of the framework rather than determining whether a particular AI system complies with specific legal or technical requirements.

A first validation approach consists of **expert review**, involving multidisciplinary specialists with complementary expertise in AI, law enforcement operations, data protection, fundamental rights, cybersecurity, and ethics. Expert assessment should evaluate whether the governance model adequately captures the principal legal, organisational and technical dimensions associated with AI-enabled intelligence-support systems while remaining sufficiently flexible to accommodate different operational contexts. Particular attention should be paid to the coherence of the governance structure, the relevance of the proposed safeguards, the appropriateness of the selected indicators and the overall usability of the framework by organisations with different levels of governance maturity.

A second validation mechanism is represented by **tabletop exercises**, through which representative operational scenarios can be analysed in a structured but controlled environment. Tabletop exercises enable practitioners to simulate governance decisions, evaluate escalation procedures, assess the interaction between different organisational actors, and verify whether the proposed safeguards support informed decision-making under realistic operational conditions. This approach is particularly valuable in the FCT domain, where many governance challenges arise from organisational processes and interdependencies.

The model should also be refined through **structured practitioners' feedback** collected from the organisations expected to apply it in practice. Investigators, intelligence analysts, operational supervisors, data protection officers, and AI developers each interact with governance processes from different perspectives and therefore provide complementary insights into the practical usability of the framework. Structured feedback should not be limited to identifying missing safeguards but should also examine whether the governance process remains proportionate, sufficiently clear for operational use and capable of supporting decision-making without introducing unnecessary administrative complexity. This iterative process strengthens both the practical applicability and the long-term sustainability of the methodology.

An additional source of validation can be obtained through **comparison with the findings emerging from the ENACT Structured Knowledge Base (SKB)** and other project outputs. Although the governance model is intended to remain technology-neutral, comparison with empirical evidence collected across representative operational contexts allows verification that the identified safeguards, indicators and governance concerns adequately reflect the risks observed within contemporary FCT research and innovation activities. Such comparison also contributes to identifying emerging governance challenges that may require future refinement of the methodology as AI technologies and investigative practices continue to evolve.

Finally, the governance model should be **tested against representative case studies** reflecting different categories of AI-enabled intelligence-support technologies, such as those presented earlier in this section. Applying the framework across diverse operational scenarios enables verification that the proposed governance structure remains sufficiently robust to accommodate different technological architectures, data environments, and investigative objectives while maintaining a consistent methodology for assessing safeguards and documenting evidence. Successful application across heterogeneous use cases would provide an important indication of the framework's generalisability without compromising its flexibility.

Importantly, none of these validation activities should be interpreted as providing formal certification of an AI system or demonstrating legal compliance with the AI Act, the GDPR, the LED or other applicable regulatory instruments. Compliance assessment remains the responsibility of the relevant legal and regulatory mechanisms applicable to each operational context. Instead, validation aims to determine whether the proposed governance model constitutes a **complete, coherent, and practically useful governance instrument** capable of supporting organisations in implementing accountable, evidence-based, and rights-respecting AI governance.

This distinction is particularly significant within the FCT R&I ecosystem, where AI technologies continue to evolve rapidly, and governance requirements must remain adaptable to changing operational needs, technological developments, and regulatory expectations. By focusing on the effectiveness of governance processes, the proposed methodology encourages continuous organisational learning, supports iterative improvement, and facilitates the gradual development of governance maturity across research projects and operational environments. In this respect, the framework should be regarded as a living governance instrument capable of evolving alongside both technological innovation and the European regulatory landscape.

RECOMMENDATIONS FOR IMPLEMENTATION AND PRACTICAL UPTAKE

The governance framework developed in this report is intended to support the responsible design, deployment and continuous improvement of AI-enabled intelligence-support technologies across the European FCT R&I ecosystem. While the principles and safeguards discussed in the previous sections provide a common governance structure, their effective implementation ultimately depends on the coordinated actions of multiple stakeholders exercising distinct legal, technical, and organisational responsibilities. Accordingly, successful AI governance cannot be achieved through technical measures alone but requires the integration of governance arrangements throughout the entire lifecycle of AI systems.

An important consideration emerging from this report is that **responsibility for implementing safeguards is distributed across different actors**. Certain measures are inherently technical and therefore fall primarily within the responsibilities of developers and technology providers, including explainability functionalities, audit logging capabilities, confidence information, robustness testing and mechanisms supporting meaningful human oversight. Other safeguards, however, cannot be embedded exclusively within the technology itself because they depend upon organisational or legal decisions taken by deploying organisations. These include the definition of lawful purposes, retention policies, allocation of accountability, authorisation procedures, escalation mechanisms, staff training, operational monitoring, complaint handling, audit planning, and the periodic review of governance arrangements following material organisational or technological changes.

Recognising this distinction is essential to avoid unrealistic expectations regarding the capacity of AI systems to guarantee legal compliance independently of the organisations responsible for their use. This distinction is particularly relevant in EU-funded research and innovation projects, where technology developers, research organisations, Law Enforcement Authorities and other stakeholders collaborate throughout the design, testing and validation phases. Such collaborative environments provide valuable opportunities to embed legal, ethical and operational safeguards by design, while also requiring a clear allocation of roles and responsibilities to facilitate the transition from research prototypes to trustworthy operational deployment.

The proposed governance model therefore supports a lifecycle approach structured around three complementary implementation phases. **Before deployment**, organisations should establish the legal qualification of the intended processing activity, perform an initial risk screening, conduct the appropriate DPIA and, where relevant, a FRIA, validate system performance against representative operational scenarios, define governance responsibilities, and ensure that all personnel involved possess the necessary operational, legal and technical competencies. **During operational use**, governance should focus on the continuous monitoring of key performance indicators, periodic auditing of safeguard effectiveness, review of incidents and operator interventions, representative sampling of operational decisions, monitoring of model drift and performance degradation, and the effective management of complaints and redress mechanisms.

Following any material change affecting the AI system, its operational environment or the applicable legal framework, organisations should perform a renewed assessment of risks, update technical and governance documentation, review safeguards, validate modified system behaviour, and provide additional training where necessary. In this manner, governance evolves as a continuous organisational process, rather than a one-off compliance exercise conducted immediately before deployment.

The practical uptake of the framework should likewise be supported through governance artefacts that facilitate its adoption by organisations with different levels of technical and organisational maturity. The **Safeguards Effectiveness Matrix** provides the primary operational instrument for documenting safeguards, responsibilities, evidence, and monitoring indicators. This can be complemented by a concise governance model summarising the principal governance layers, a practitioner-oriented checklist supporting operational reviews, a decision tree assisting organisations in selecting the appropriate governance measures for specific use cases, and short policy-oriented guidance capable of informing procurement strategies, research programmes, and institutional governance policies. Together, these outputs translate the conceptual framework into practical tools that can be readily integrated within research projects, procurement procedures and operational governance activities.

Table 7 summarises the principal recommendations emerging from the governance framework according to the main stakeholder groups participating in the European FCT research and innovation ecosystem.

Table 7: Recommendations by stakeholder group

STAKEHOLDER	KEY RECOMMENDATIONS
EU policymakers and funders	Promote harmonised governance methodologies; require lifecycle governance and fundamental rights considerations within funded projects; encourage interoperability and standardisation of governance practices.
Law Enforcement Authorities (Deployers)	Establish multidisciplinary governance structures; conduct legal and risk assessments before deployment; implement meaningful human oversight; monitor safeguard effectiveness; review governance after material changes.
Researchers and project consortia	Integrate legal, ethical and operational requirements from the design phase; adopt evidence-based governance methodologies; validate governance tools with practitioners throughout the project lifecycle.
Technology providers and developers	Design systems supporting explainability, auditability, logging, robustness and effective human oversight; provide comprehensive technical documentation; facilitate continuous monitoring and independent evaluation.
Oversight bodies	Encourage evidence-based auditing approaches; promote harmonised interpretation of governance requirements; support continuous dialogue between regulators, practitioners and researchers; disseminate good governance practices across sectors.

All the abovementioned recommendations reinforce the central conclusion of this report: trustworthy AI within the FCT domain depends not on the existence of individual safeguards, but on the establishment of an integrated governance ecosystem in which technical capabilities, organisational procedures, and legal accountability operate in a coherent and mutually reinforcing manner. By combining lifecycle governance, multidisciplinary oversight and continuous evidence-based evaluation, the proposed framework provides a practical foundation for supporting innovation while safeguarding fundamental rights, strengthening institutional accountability and promoting public trust in the responsible use of AI-enabled intelligence-support technologies.

49 CONCLUSIONS

AI is rapidly transforming the analytical capabilities available to LEAs by enabling the processing of large and heterogeneous datasets, supporting intelligence generation, and assisting operational decision-making across increasingly complex investigative environments. Within the FCT domain, these technological developments offer significant opportunities to strengthen investigative efficiency, improve situational awareness, and facilitate cross-border cooperation. At the same time, they introduce new legal, ethical and societal challenges that cannot be addressed solely through technical innovation or regulatory compliance. The responsible use of AI-enabled intelligence-support technologies requires governance approaches capable of integrating technological performance with accountability, transparency, fundamental rights protection, and institutional legitimacy throughout the entire system lifecycle.

This report has addressed these challenges by proposing a structured governance framework that combines legal analysis, ethical principles, operational safeguards and evidence-based monitoring into a coherent methodology applicable across different AI-enabled intelligence-support scenarios, as illustrated through the representative operational scenarios examined throughout the report: OSINT and social-media intelligence, multilingual and multimodal content analysis, and graph-based intelligence and relationship detection. The report has adopted a lifecycle perspective in which governance accompanies the design, development, deployment, operation and continuous improvement of AI systems. This approach reflects the growing convergence between the European regulatory framework, contemporary AI governance models and the practical needs of organisations operating in complex security environments.

A central contribution of the report is the recognition that trustworthy AI cannot be achieved through compliance with individual legal obligations or through the implementation of isolated technical safeguards. Instead, governance should be understood as a multidimensional organisational capability requiring the interaction of legal qualification, risk management, meaningful human oversight, organisational accountability, and continuous monitoring. Throughout the report, particular emphasis has been placed on the complementary nature of these governance dimensions, recognising that technical safeguards are effective only when supported by appropriate organisational structures, competent personnel, and clearly defined institutional responsibilities.

Building upon this perspective, the report has introduced several practical governance instruments intended to facilitate the operational implementation of trustworthy AI within the FCT research and innovation ecosystem. The four-layer safeguards model provides a structured representation of the different governance dimensions that should accompany AI-enabled intelligence-support technologies. The analysis of meaningful human oversight demonstrates that effective supervision depends not only on technical functionalities, but also on organisational design, operational procedures and human expertise.

The Safeguards Effectiveness Matrix extends this approach by providing a practical mechanism for linking safeguards, risks, responsibilities, evidence and performance indicators, thereby enabling organisations to evaluate not merely the existence of governance measures but their continuing effectiveness over time.

The application of the proposed methodology to representative operational scenarios has further demonstrated that governance challenges remain remarkably consistent across different technological solutions. Whether analysing publicly available information through OSINT, processing multilingual and multimodal content or identifying relationships through graph-based intelligence, similar governance requirements emerge regarding data quality, proportionality, explainability, accountability, human oversight, and continuous organisational learning. This confirms that governance frameworks should remain technology-neutral while providing sufficient principles to accommodate rapidly evolving analytical capabilities. The implications of this work extend beyond individual AI applications and are directly relevant to the broader European FCT R&I ecosystem. As AI technologies become increasingly integrated into collaborative research projects, procurement activities, and operational environments, governance must similarly evolve from isolated compliance exercises towards continuous organisational processes supported by multidisciplinary expertise. Integrating governance considerations into research design, project management, procurement strategies, deployment planning and post-deployment monitoring contributes both to legal compliance and to strengthening public trust, institutional accountability, and long-term sustainability of AI-enabled innovation.

An equally important conclusion emerging from this report concerns the distribution of responsibilities among stakeholders. The implementation of trustworthy AI cannot be delegated exclusively to technology providers, nor can it rely solely upon organisational procedures adopted by deployers. Effective governance requires coordinated action by developers, law enforcement authorities, controllers, procurement bodies, oversight authorities, policymakers, and research organisations, each exercising responsibilities that reflect their respective legal mandates and operational roles. Recognising these differentiated responsibilities is essential to ensure that governance measures remain both effective and proportionate throughout the AI lifecycle.

The proposed framework should therefore be regarded as a practical governance methodology rather than as a certification scheme or a static compliance checklist. Its objective is to support informed decision-making by enabling organisations to identify governance requirements, document evidence, monitor safeguard effectiveness, and continuously improve organisational practices in response to evolving technologies, operational needs and regulatory developments. Future refinement of the methodology should continue through multidisciplinary validation, empirical application within representative operational environments and alignment with emerging European standards and guidance concerning trustworthy AI and fundamental rights protection.

Finally, this report contributes to the ongoing European discussion concerning the responsible adoption of AI within the security domain by demonstrating that innovation and fundamental rights protection should not be viewed as competing objectives. On the contrary, effective governance constitutes a strategic enabler of sustainable innovation. AI systems that are legally sound, ethically robust, operationally transparent, and organisationally accountable are more likely to achieve public legitimacy, facilitate cross-border cooperation, and support the long-term effectiveness of European R&I in FCT. By integrating legal analysis, ethical reflection, operational governance and evidence-based evaluation within a single methodological framework, this report seeks to provide a practical contribution towards the development of trustworthy, human-centred, and accountable AI-enabled intelligence-support technologies capable of serving both operational effectiveness and the rule of law.

BIBLIOGRAPHY

ALIGNER Project. (2024). *Fundamental Rights Impact Assessment (FRIA) methodology for AI in law enforcement.* https://aligner-project.eu/

Article 29 Data Protection Working Party. (2017). *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 (WP248 rev.01).* https://ec.europa.eu/newsroom/article29/items/611236

Council of Europe. (1950). *European Convention on Human Rights.* https://www.echr.coe.int/

Council of Europe. (2019). *Guidelines on Artificial Intelligence and Data Protection (T-PD(2019)01).* https://rm.coe.int/

Commission Nationale de l'Informatique et des Libertés (CNIL). (2023). *Self-assessment guide for artificial intelligence (AI) systems.* https://www.cnil.fr/

European Commission. (2019). *Ethics guidelines for trustworthy AI.* High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/

European Commission. (2024). *Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act).* https://digital-strategy.ec.europa.eu/

European Commission. (2024). *Draft Commission guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689 (AI Act).* https://digital-strategy.ec.europa.eu/

European Data Protection Board. (2018). *Opinion 12/2018 on a draft list of the competent supervisory authority of Belgium regarding the processing operations subject to the requirement of a data protection impact assessment.*

European Data Protection Board. (2020). *Guidelines 07/2020 on the concepts of controller and processor in the GDPR.* https://edpb.europa.eu/

European Data Protection Board. (2026). *Data Protection Impact Assessment (DPIA) template.* https://www.edpb.europa.eu/

European Data Protection Supervisor. (2016). *An ethical approach to fundamental rights.* Publications Office of the European Union.

European Data Protection Supervisor. (2024). *Checklist on human intervention in AI-supported decision-making.* https://www.edps.europa.eu/

European Data Protection Supervisor. (2024). *Guidelines on generative AI: Embracing opportunities, protecting people.* https://www.edps.europa.eu/

European Data Protection Supervisor. (2024). *TechDispatch: Human oversight.* https://www.edps.europa.eu/

European Parliament & Council of the European Union. (2012). *Charter of Fundamental Rights of the European Union (2012/C 326/02).* https://eur-lex.europa.eu/

European Parliament & Council of the European Union. (2016). *Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties.* https://eur-lex.europa.eu/

European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation).* https://eur-lex.europa.eu/

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* https://eur-lex.europa.eu/

Europol Innovation Lab. (2025). *Assessing technology in law enforcement: A method for ethical decision-making* (Revised ed.). Europol. https://www.europol.europa.eu/

High-Level Expert Group on Artificial Intelligence. (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI).* European Commission. https://digital-strategy.ec.europa.eu/

Information Commissioner's Office. (2022). *How do we ensure fairness in AI?* https://ico.org.uk/

International Organization for Standardization. (2023). *ISO/IEC 23894:2023—Information technology—Artificial intelligence—Guidance on risk management.*

International Organization for Standardization. (2023). *ISO/IEC 42001:2023—Information technology—Artificial intelligence—Management system.*

Mantelero, A. (2024). *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template.* Computer Law & Security Review, 54, 106019.

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* U.S. Department of Commerce. https://www.nist.gov/

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449).* https://legalinstruments.oecd.org/

Spanish Data Protection Agency. (2021). *Audit requirements for personal data processing activities involving artificial intelligence.* https://www.aepd.es/

Sterz, A., et al. (2024). *On the quest for effectiveness in human oversight: Interdisciplinary perspectives. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24).* Association for Computing Machinery.

U.S.–EU Trade and Technology Council. (2022). *Joint roadmap on evaluation and measurement tools for trustworthy AI and risk management.*https://www.trade.gov/

Wachter, S., Mittelstadt, B., & Russell, C. (2024). *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law & Security Review.* Advance online publication. https://doi.org/10.2139/ssrn.3547922

ENACT.
European Network Against Crime and Terrorism

@enact-network

enact-eu.net

Scan the QR code to visit ENACT online and read this report in digital format.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.