دار نشر جامعة لوسيل
Lusail University Press

RESEARCH ARTICLE

Identifying and Outlining the Governing Principles on the Use of Artificial Intelligence in Anti-Corruption Laws, Policies, and Institutions

Qerim Qerimi

Qerim Qerimi
University of Prishtina, Prishtina 10000, Kosovo
qerim.qerimi@uni-pr.edu
ORCID ID: https://orcid.org/0000-0003-4296-7650

Submitted: 03 November 2025
Accepted: 14 February 2026

https://doi.org/10.70139/rolacc.2026.1.1

Cite this article as: Qerimi, Q. Identifying and Outlining the Governing Principles on the Use of Artificial Intelligence in Anti-Corruption Laws, Policies, and Institutions, Rule of Law and Anti-Corruption Center Journal, 2026:1, https://doi.org/10.70139/rolacc.2026.1.1

Abstract

The increasing deployment of Artificial Intelligence (AI) in public governance raises critical questions about its role, legitimacy, benefits and risks in discrete domains, including anti-corruption. While AI offers means and ways for detecting patterns of fraud, analyzing complex data, its use in anti-corruption laws, policies, and institutions (in particular those specialized anti-corruption bodies), remains hitherto under-regulated and insufficiently theorized. This article seeks to advance a normative framework of fundamental principles to guide the responsible use of AI in anti-corruption contexts. The core referential or inspirational basis comes out of the underlying principles enshrined in the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, and the Venice Commission's standards, including its Rule of Law Checklist, as those standards are already applied in contexts pertinent to good governance, such as the "Interpretative Declaration of the Code of Good Governance in Electoral Matters as Concerns Digital Technologies and Artificial Intelligence". Adapting key principles such as legality, transparency, accountability, non-discrimination, human oversight, and proportionality, the article outlines a set of safeguards necessary to preserve the rule of law, institutional independence, and human rights and fundamental freedoms in the face of expanding AI applications. It posits that without a clear regulatory and ethical framework, the use of AI risks reinforcing existing biases, undermining due process, and enabling politically motivated manipulation of anti-corruption mechanisms. The proposed framework emphasizes the necessity of legal clarity, human-centric oversight, independent evaluation, and public transparency. It further underscores the importance of inclusive design, data protection, and international cooperation in shaping AI use that is both effective and legitimate. By grounding AI deployment in consolidated democratic values and institutional integrity, the framework of principles seeks to ensure that AI serves as a tool to strengthen rather than compromise the global fight against corruption. The result of this proposal could be a Draft Declaration of Core Principles on the Use of Artificial Intelligence in Anti-Corruption Laws, Policies, and Institutions. This article further illustrates the proposed principles through selected case vignettes and situates the analysis within a global anti-corruption framework, including the practice and implementation mechanisms of the United Nations Convention against Corruption.

Keywords: Artificial intelligence; anti-corruption governance; rule of law and human rights; accountability and transparency; ethical AI; AI regulation

1. INTRODUCTION

The deployment of Artificial Intelligence (AI) across public governance frameworks and structures is reshaping how states and state authorities prevent, detect, and respond to corruption. Automated systems increasingly support data analytics for identifying procurement irregularities, mapping conflicts of interest, and flagging suspicious transactions.[1] Some states (e.g., Albania) have even begun appointing AI ministers to oversee procurement processes and battle corruption.[2] Yet the legal and ethical frameworks governing these uses remain largely underdeveloped. While AI has potential for efficiency and predictive accuracy, it also introduces bias and accountability deficits that challenge the foundational principles of the rule of law.[3]

Specifically, specialized anti-corruption institutions, such as anti-corruption agencies, integrity commissions or judicial and prosecutorial bodies, pervade and shape the central legal and democratic landscape of anti-corruption efforts. Their legitimacy depends on independence, due process, and public trust. The integration of AI into their operations raises complex normative questions. Some of the fundamental questions include: What normative principles should guide and govern the integration of AI? How should automated systems be regulated to preserve human oversight, independence, and fairness (both procedural and substantive)? And what guardrails are needed to guarantee that AI enhances, rather than undermines, legality, transparency, accountability, and equality before the law?

This article advances the argument that the use of AI in anti-corruption contexts must be grounded in a normative framework of governing principles derived from, or drawing on, existing international legal standards, particularly those of the Council of Europe's (CoE) Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law[4] and the Venice Commission's Rule of Law Checklist.[5] These instruments provide the foundational normative infrastructure necessary to guide and align AI governance in the arena of anti-corruption with fundamental rule of law principles and international best practice.

This article seeks to serve at least the following three objectives. First, it delimits the normative standards of AI regulation as those standards relate to anti-corruption governance. Second, it outlines a comprehensive principled approach to the responsible use of AI by anti-corruption institutions, underscoring the indispensable value of the composite pillars of legality, transparency, accountability, non-discrimination, and human oversight. Third, it proposes a Draft Declaration of Core Principles on the Use of AI in Anti-Corruption Contexts, annexed hereto, and intended to guide both state practice and academic discourse and policy debate.

The article is organized in the following manner. Section 2 theorizes the intersection of AI, corruption, and the rule of law. Section 3 examines the normative foundations for AI governance. It focuses on the CoE and complementary global standards. Section 4 constructs and elaborates a framework to guide the discussion on, and regulation of, the AI's use in the pertinent anti-corruption context. Section 5 introduces the proposed Draft Declaration, followed by concluding reflections in Section 6.

This article adopts a normative doctrinal methodology grounded in public international law, comparative constitutional law, and rule of law scholarship. It identifies, interprets, and operationalizes governing principles applicable to the use of AI in anti-corruption laws, policies, and institutions by drawing on binding and non-binding international instruments and established rule of law standards.

The selection of normative instruments is guided by three criteria. First, priority is given to instruments that articulate binding or authoritative standards relating to democracy, human rights, and the rule of law. Second, preference is accorded to frameworks that apply to public authorities exercising investigative, prosecutorial, or adjudicative functions. Third, the instruments selected demonstrate either explicit relevance to AI governance or functional applicability to algorithmic decision-making. On this basis, the CoE's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, the European Union (EU) Artificial Intelligence Act, and the Venice Commission's Rule of Law Checklist, dealing

1 See generally: Organisation for Economic Co-operation and Development, *Using Artificial Intelligence to Combat Corruption in Public Procurement* (2021).

2 See, e.g., Guy Delauney, *World's First AI Minister Will Eliminate Corruption, Says Albania's PM*, *BBC News* (Sept. 12, 2025), https://www.bbc.com/news/articles/cm2znzgwj3xo; *Albania Appoints AI Bot Minister to Fight Corruption in World First*, Al Jazeera (Sept. 12, 2025), https://www.aljazeera.com/news/2025/9/12/albania-appoints-ai-bot-minister-to-fight-corruption-in-world-first; Tharin Pillay, *The World's First AI-Powered Minister Tests the Future of Government*, TIME (Sept. 12, 2025), https://time.com/7324934/albania-ai-minister-die.

3 Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147, 1150-52 (2017).

4 Council of Eur., Framework Convention on Artificial Intelligence and Human Rights, *Democracy and the Rule of Law* (May 2024) [hereinafter Framework Convention on AI or CoE Framework Convention on AI].

5 Eur. Comm'n for Democracy Through Law (Venice Comm'n), *Rule of Law Checklist*, CDL-AD(2016)007 (Mar. 18, 2016) [hereinafter Rule of Law Checklist].

specifically with matters concerning the governance of AI, constitute the primary normative reference points. These instruments are complemented by global frameworks that are rather more significant on the side of subject-matter than the scope and substance of AI-regulation, notably the United Nations Convention against Corruption (UNCAC), United Nations Office on Drugs and Crime (UNODC) guidance, and Organisation for Economic Cooperation and Development (OECD) recommendations, which together reflect an emerging convergence around human-centered, risk-based AI governance.

The interpretative approach is functional and purposive. Norms and standards are analyzed not in isolation, but in light of the institutional realities of anti-corruption governance, including the exercise of discretion, institutional and financial independence, due process guarantees, and public accountability. The interpretation employed is thus contextual, drawing on existing jurisprudence, institutional practice, and scholarly literature to adapt rule of law principles to the specific risks posed by AI-enabled decision-making.

The underlying methodology necessarily entails limitations. The article does not seek to offer quantitative evaluation of AI effectiveness, nor does it assess proprietary systems deployed in specific jurisdictions. Instead, it aims to provide a transferable normative framework capable of guiding legislators, governmental institutions or executive agencies, and courts across diverse legal systems and cultures. The proposed Draft Declaration should in this context be understood as a normative and governance-oriented instrument rather than a technical or sector-specific regulatory document.

2. AI, THE RULE OF LAW, AND ANTI-CORRUPTION EFFORTS

The global expansion of AI in governance coincides with renewed international attention to integrity, accountability, and anti-corruption.[6] AI systems are increasingly used to enhance transparency, strengthen audit capacity, and detect defects in financial and procurement data.[7] These applications commonly involve machine learning, natural language processing, and predictive analytics to identify suspicious transactions, evaluate corruption risks, and support monitoring of public spending.

In theory, these applications can mitigate corruption risks by enabling data-driven decision-making and reducing human discretion. However, without appropriate legal and ethical frameworks and safeguards, AI may produce the opposite effect, such as embedding opacity, reinforcing discriminatory biases, and facilitating forms of legal uncertainty.[8] In this context, appropriate human oversight becomes central. This is demonstrated also by the scholarly consensus and ensuing understanding that human oversight is essential not only for ensuring accountability but also for promoting and supporting the broader goals of societal betterment, well-being, and justice.[9]

2.1. Defining and measuring the rule of law in the era of AI

In academic discourse and pertinent practice, the rule of law is understood to encompass both procedural and substantive guarantees. It is a foundational principle of governance mandating that all persons, legal or natural, are bound by and accountable to properly promulgated laws. These laws should conform to internationally recognized rights and freedoms.[10]

The Rule of Law Checklist of the Venice Commission provides a practical tool for assessing and measuring compliance, conceiving the rule of law as "a concept of universal validity"[11]. When applied to AI governance, the Venice Commission's rule of law elements yield several normative imperatives. First, legality demands that AI systems used in public administration or anti-corruption contexts are clearly authorized by law or legal norms. Second, legal certainty demands that laws or legal norms authorizing the use of AI systems are accessible and foreseeable. Third, accountability requires identifiable responsibility for AI-assisted decisions, so as to avoid arbitrariness and abuse of power (*détournement de pouvoir*). Fourth, non-discrimination ensures that the benefits of AI are pursued

6 Transparency Int'l, *The Corruption Risks of Artificial Intelligence* (2022).

7 OECD, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions* (2025).

8 Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J.L. & TECH. 347 (2021).

9 See, e.g., Reuben Binns, *Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making*, 16 REGUL. & GOVERNANCE 197 (2022); Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429 (2023); Siavosh Sahebi & Paul Formosa, *Artificial Intelligence and Global Justice*, 35 MINDS & MACHS. (forthcoming 2025); Daniel J. Solove & Hideyuki Matsumi, *AI, Algorithms, and Awful Humans*, 92 FORDHAM L. REV. 1923, 1923 (2024); Ben Wagner, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems*, 11 POL'Y & INTERNET 104, 104 (2019).

10 U.N. Secretary-General, *The Rule of Law and Transitional Justice in Conflict and Post-Conflict Societies*, ¶ 6, U.N. Doc. S/2004/616 (Aug. 23, 2004).

11 Venice Comm'n, Rule of Law Checklist, *supra* note 5. See also: Qerim Qerimi, *Operationalizing and Measuring Rule of Law in an Internationalized Transitional Context: The Virtue of Venice Commission's Rule of Law Checklist*, 13 Law & Dev. Rev. 59, 59 (2020). The composite elements of the rule of law include: legality, legal certainty, prevention of abuse (misuse) of power, equality before the law and non-discrimination, and access to justice.

on the basis of equality and without infringing fundamental rights. Fifth, transparency and explainability become key procedural components of due process, enabling affected persons to understand and challenge AI-driven outcomes and, therefore, enabling or realizing access to justice.

AI thus constitutes both a test and a tool for the rule of law. It can strengthen good governance by reducing arbitrariness, but it can also erode legal certainty if decisions are made through non-transparent algorithms shielded from scrutiny.[12] Thus, as algorithmic regulation becomes embedded in legislative, judicial, and other legal and institutional practices, it must not only align with the rule of law but also embody its fundamental precepts.[13]

2.2. Corruption, automation, and government risks

Corruption, conceived widely as the "abuse of entrusted power"[14] or "of public office or private office for personal gain",[15] remains a permanent and structural challenge to good governance. The UNCAC refrains from defining corruption. Rather, it adopts a broad approach, identifying a range of corrupt practices and establishing the most authoritative global legal framework for anti-corruption measures.[16]

AI systems can assist anti-corruption bodies by identifying risk patterns across procurement data, monitoring conflicts of interest, or predicting potential fraud.[17] Yet automation introduces new challenges and vulnerabilities.

First, algorithmic decision-making may replicate or amplify existing societal biases if training data or design reflect discriminatory patterns.[18] Second, AI systems' "black-box" decision-making can undermine procedural fairness and accountability, particularly when used in investigative or prosecutorial contexts.[19] Third, the concentration of technological control, which is when private vendors develop or maintain AI systems used by public authorities, creates dependencies that may compromise institutional independence.[20] Finally, the strategic misuse of AI (for instance, for surveillance or such other purposes as targeting under the guise of anti-corruption) can become an instrument of political manipulation, contradicting both democracy and the rule of law.[21]

These risks underscore the necessity of grounding AI deployment within an explicit and consolidated legal framework. In this connection, the rule of law must remain the foundational framework of AI governance in any context, including anti-corruption. The following sections therefore explore how existing legal standards, in particular the Framework Convention on AI and the Venice Commission's instruments and standards, can inform and guide principled AI use in anti-corruption contexts.

3. NORMATIVE FOUNDATIONS FOR AI GOVERNANCE

The normative architecture for governing AI is constantly evolving, with European institutions leading the way in shaping its key principles and frameworks. While numerous soft-law instruments address ethical principles, and select hard-law instruments concern data protection, few are directly aimed at governing the intersection between the rule of law and AI, and even less so their combined application in anti-corruption contexts.

The Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law represents a decisive advancement in establishing binding international obligations, while complementary guidance from the Venice Commission, United Nations Convention against Corruption/UNODC, OECD, and the European Union AI Act contributes to a multi-layered normative infrastructure.

3.1. The CoE's Framework Convention on AI

In May 2024, the CoE adopted the world's first international instrument on AI and Human Rights, Democracy and the Rule of Law.[22]

12 Mireille Hildebrandt, *Algorithmic Regulation and the Rule of Law*, 376 PHIL. TRANSACTIONS ROYAL SOC'Y 1, 1 (2018)

13 *Id.* at 9.

14 European Commission, *Anti-Corruption*, https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/democracy-eu-citizenship-anti-corruption/anti-corruption_en (last visited 6 July 2026).

15 Organisation for Economic Co-Operation and Development, *Corruption: A Glossary of International Standards in Criminal Law* 22 (OECD Publishing, 2008).

16 United Nations Convention against Corruption, Oct. 31, 2003, 2349 U.N.T.S. 41.

17 OECD, *Analytics for Integrity: Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments* (2019).

18 Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, 41 COMPUTER L. & SEC. REV. 1, 1 (2021).

19 Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1085 (2018).

20 David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey & Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, ADMIN. CONF. U.S. (2020).

21 Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021); Jürgen Habermas, *On the Internal Relation Between the Rule of Law and Democracy, in* THE INCLUSION OF THE OTHER: STUDIES IN POLITICAL THEORY 253 (Ciaran Cronin & Pablo De Greiff eds., Polity 2015).

22 Framework Convention on AI, *supra* note 4.

The Framework Convention introduces legally binding obligations.[23] Of special significance, it mandates States Parties to establish independent oversight mechanisms in order to ensure that these obligations are enforced throughout the AI lifecycle,[24] in relation to both the activities and content that is generated by AI systems.[25]

Human oversight is thus a fundamental legal building block that should define the new regulatory frameworks aimed at addressing the relationship with AI systems and be followed by transparent governance and procedural guardrails. It is worth emphasizing that the suggestion that oversight intensity should scale with risk[26] is clearly reflected in the treaty's risk-based approach, as outlined in Article 16.[27] This approach aligns with best practices in assigning oversight modalities proportionate to an AI system's calculated risk and potential impact. Accordingly, the Framework Convention's human oversight mandate establishes a "graduated and differentiated"[28] oversight regime, recognizing that states will need to make explicit choices regarding the form, depth, and criteria for oversight based on the probability and severity of negative impact.[29]

While the Framework Convention represents a significant legal advance by qualifying and codifying human oversight as both a core legal principle and a state obligation, meaningful human oversight is not automatic, as much depends on subsequent practice. It requires deliberate institutional design, adequate human and technical capacity, clear delineation of roles and responsibilities, the operationalization of context-sensitive standards, transparent decision-making processes, and safeguards against oversight failure and capture. These practice-oriented considerations should guide the implementing philosophy of the Framework Convention's oversight requirements.

Transparency and procedural safeguards are other core components of the Convention's approach.[30] Individuals interacting with AI systems have the right to be informed when they are engaging with such systems rather than humans, and authorities have a corresponding obligation to ensure this. Moreover, individuals must have access to legal remedies that are effective in case of harm or unfair treatment.[31] This ensures not just technical compliance, but democratic legitimacy in how AI is deployed and used in public life. The Venice Commission has echoed this procedural safeguard in its "Interpretative Declaration of the Code of Good Practice in Electoral Matters as Concerns Digital Technologies and Artificial Intelligence", providing that whenever AI systems "are being used in elections, voters should be informed that they are interacting with such systems rather than with a human."[32]

The Convention does not duplicate existing human rights law. Rather, it complements it by performing a dual function: first, by promoting AI systems to protect and advance human rights, democracy and the rule of law; and, second, by safeguarding against the abuse or misuse of such systems to the detriment of human rights, democracy and the rule of law, which are "inherently interwoven."[33] It is explicitly designed to function alongside international and regional instruments such as the Universal Declaration of Human Rights and the European Convention on Human Rights, thereby situating AI governance within the already established international legal system.[34] Given that the Convention is open to non-European states, it offers a platform for norm diffusion and convergence on a global scale. In the European context, it is designed to operate harmoniously with the EU's AI Act, bridging human rights law with market regulation.

3.2. The Venice Commission's Rule of Law Checklist and emerging AI guidance

The Venice Commission is an independent advisory body of the CoE. It acts as a guardian of constitutional and democratic standards within the framework of the CoE. Its Rule of Law Checklist, adopted in 2016 and updated in 2025, remains a

23 Framework Convention on AI, *supra* note 4, Ch. III.

24 *Id.* art. 26 ("*Each Party shall establish or designate one or more effective mechanisms to oversee compliance with the obligations in this Convention.*").

25 *Id.* art. 8.

26 Laxmiraju Kandikatla & Branislav Radeljić, *AI and Human Oversight: A Risk-Based Framework for Alignment* (2025), https://arxiv.org/abs/2510.09090.

27 Framework Convention on AI, *supra* note 4, art. 16.

28 *Id.* art. 1(2).

29 *Id.*

30 *Id.* arts. 8 & 15.

31 *Id.* art. 15, ¶ 2.

32 Eur. Comm'n for Democracy Through Law (Venice Comm'n), *Interpretative Declaration of the Code of Good Practice in Electoral Matters as Concerns Digital Technologies and Artificial Intelligence*, CDL-AD(2024)044 (Dec. 10, 2024), ¶ 8.

33 Framework Convention on AI, *supra* note 4, pmbl.

34 *Id.*

cornerstone reference for assessing or measuring rule of law in practical terms.[35] The Checklist translates the rule of law into practice through five foundational elements, namely "legality, legal certainty, prevention of abuse (misuse) of powers, equality before the law and non-discrimination, and access to justice."[36] The Updated Rule of Law Checklist introduced checks and balances and constitutional review as independent, standalone elements.

In the emerging AI context, these principles translate into enforceable obligations of legality, accountability, and fairness. The Venice Commission has underscored that, within the sphere of electoral processes, electoral authorities bear the responsibility to ensure that the use of AI is compatible with the rule of law principles.[37]

3.3. Complementary global instruments: UNCAC/ UNODC, OECD, and the EU AI Act

While the CoE provides the most sophisticated rule of law-based approach, global and regional institutions and instruments contribute additional layers of normative guidance relevant to anti-corruption applications of AI.

3.3.1. UNCAC/UNODC

The UNCAC, which is presently ratified by 192 parties, remains the primary international treaty addressing corruption prevention, criminalization, and institutional integrity.[38] Although UNCAC predates contemporary AI regulation, its provisions on preventive measures, transparency, and public sector integrity establish a legal foundation for responsible AI use.

Article 5 of UNCAC obliges states parties to adopt and implement effective anti-corruption policies.[39] When integrated with AI governance, this obligation entails that states ensure algorithmic systems' compliance with the rule of law principles. Similarly, the requirement to create independent anti-corruption bodies implies that AI deployment within, and use by, such bodies must not compromise their operational independence or professional impartiality, so as to discharge their functions effectively and free from improper influence.[40]

The UNODC is the designated body to discharge the functions of secretariat of the Conference of the States Parties to the Convention.[41]

In terms of global practice, several jurisdictions are experimenting with digital and AI-enhanced tools to strengthen integrity and compliance with UNCAC-aligned preventive measures. In Brazil, for example, the Office of the Comptroller General (Controladoria-Geral da União – CGU) had deployed advanced procurement analytics such as the AI-powered "Alice" system to continuously analyze federal tenders and flag irregularities in bids, contracts and notices, significantly accelerating audit workflows and enhancing oversight of high-risk public spending.[42]

In the Republic of Korea, government digitalization initiatives have incorporated platforms such as the Clean Portal, a digital reporting mechanism launched by the Anti-Corruption & Civil Rights Commission to streamline corruption and public interest violation reporting and support data-driven anti-corruption policymaking.[43] In Indonesia, the Corruption Eradication Commission (Komisi Pemberantasan Korupsi – KPK) has utilized citizen-centric data collection and risk mapping tools such as the mobile app "JAGA", which aggregates local governance data to identify potential corruption risks and trigger further analysis.[44] The app lets citizens access and verify government funding data, such as school scholarships and whether the allocated funds have reached their destination, as well as submit sector-specific complaints across sectors it covers, including education, health, villages, regions, budgets, and licensing.[45]

3.3.2. OECD

The OECD Principles on Artificial Intelligence (2019, amended in 2024), endorsed by nearly fifty countries and the EU, constitute a leading soft-law instrument underscoring human-centered values and accountability in AI design and use.[46] Principle 1.2 affirms that AI systems should respect

35 Venice Comm'n, Rule of Law Checklist, *supra* note 5.

36 *Id.*

37 Interpretative Declaration of the Code of Good Practice in Electoral Matters as Concerns Digital Technologies and Artificial Intelligence, *supra* note 32, ¶ 64.

38 United Nations Convention against Corruption, *supra* note 16.

39 *Id.* art. 5.

40 *Id.* art. 6.

41 *Id.* pmbl.

42 OECD, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, *supra* note 7, at 207-08.

43 UNDP Seoul Pol'y Ctr., *Introduction to Korea's Clean Portal: Digital Corruption and Public Interest Violation Reporting Mechanism* (2024).

44 U.K. Foreign, Commonwealth & Dev. Off., *International Approaches to Recording Corruption and Monitoring Intervention Outcomes at the National Level* (2025).

45 *Id.*

46 OECD, *Council Recommendation on Artificial Intelligence*, OECD/LEGAL/0449 (May 22, 2019).

the rule of law, human rights and democratic values, including fairness and privacy.[47] And Principle 1.3 calls for transparency and explainability, which is to make sure that people comprehend AI-based results or outcomes.[48] In parallel, the OECD Recommendation on Public Integrity (2017) calls for risk management frameworks that anticipate emerging integrity threats, including those posed by digital technologies.[49]

These OECD instruments, albeit non-binding, reinforce a shared expectation that public institutions adopt ethical, transparent, and accountable AI governance mechanisms. For specialized anti-corruption agencies or bodies, this includes documenting algorithmic decision-making processes, conducting bias and impact assessments, and ensuring mechanisms for redress.

3.3.3. The EU AI Act

The EU AI Act is adopted in 2024. It establishes the world's first all-encompassing regulatory system for AI.[50] The Act adopts a risk-based approach. It classifies AI systems into three categories, namely (1) prohibited, (2) high-risk, and (3) limited risk.

High-risk systems, which are in principle those systems used in law enforcement, border control, and administration of justice and democratic processes, are subject to strict requirements for data governance, human oversight, transparency, and accountability.[51]

As noted in the Act, "AI systems intended for the administration of justice and democratic processes should be classified as high-risk", given their potential for exerting serious impact on democracy, the rule of law, fundamental freedoms, and the rights to a fair trial and an effective remedy.[52] It is equally important to note that, within the realm of judicial decision-making, the ultimate responsibility for decisions must rest with humans.[53]

Human oversight is a central requirement of the EU's AI Act. However, in real-life operational terms, human oversight under the AI Act must be understood as more than a formal legal obligation, as it is essentially a governance mechanism that embeds human agency, responsibility and judgment into AI systems' design, deployment, and operation. Striking the right operational balance is therefore the key to materializing the AI Act's ambition. As Sterz and others argue, oversight must be effective and that oversight will only be meaningful if the human overseer has real (a) causal power over the system, (b) epistemic access (access to relevant information), (c) self-control (i.e., not simply deferring to automation), and (d) fitting or appropriate intentions.[54] Aspects of this understanding are indeed echoed in the AI Act's Article 14, which requires that high-risk AI systems must be "designed and developed in such a way … that they can be effectively overseen by natural persons during the period in which they are in use."[55]

Although this Act primarily aims to safeguard the EU internal market, it has direct and critical relevance for anti-corruption governance. Systems used for detecting fraud, evaluating risk, or supporting prosecutorial or judicial decision-making fall directly within its high-risk category, thereby triggering enhanced obligations for risk management, human oversight, and documentation.[56] Importantly, the EU's AI Act operates in complementarity with the Framework Convention on AI, offering a regulatory counterpart to its normative commitments and international cooperation in the realization of the purpose of the Convention.[57]

Taken together, the Framework Convention, Venice Commission instruments, UNODC and OECD guidance, and the EU's AI Act collectively establish a multi-level normative framework. This framework converges on a set of core principles that can guide responsible use of AI. Building on this common core, the following Section 4 develops a coherent normative framework designed or intended specifically for anti-corruption efforts.

4. DEVELOPING A NORMATIVE FRAMEWORK FOR AI IN ANTI-CORRUPTION LAWS, POLICIES, AND INSTITUTIONS

Building on the preceding precepts and analysis, this section seeks to advance a normative framework for the accountable and legitimate operation of AI systems in the anti-corruption context. This framework is grounded in the human rights

47 OECD, *Council Recommendation on Artificial Intelligence*, *supra* note 46, ¶ 1.2.

48 *Id.* ¶ 1.3.

49 OECD, *Council Recommendation on Public Integrity*, OECD/LEGAL/0435 (Jan. 26, 2017).

50 Regulation 2024/1689 of the European Parliament and of the Council of 12 June 2024 on Laying Down Harmonised Rules on Artificial Intelligence (AI Act), 2024 O.J. (L 1689).

51 *Id.* arts. 6-9.

52 *Id.* ¶ 61.

53 *Id.*

54 Sarah Sterz, Kevin Baum, Sebastin Biewer, Holger Hermanns, Anne LauberRönsberg, Philip Meinel & Markus Langer, *On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives* (2024), https://arxiv.org/abs/2404.04059.

55 AI Act, *supra* note 50, art. 14.

56 *Id.* Annex III(6).

57 Framework Convention on AI, *supra* note 4, art. 25.

precepts, principles of democratic accountability and the rule of law, as articulated in the Framework Convention on AI, the Venice Commission's Rule of Law Checklist, and complementary international instruments. The proposed framework translates these general principles into specific operational standards for the design, deployment, use, and oversight of AI systems in anti-corruption contexts.

4.1. Legality and legal certainty

At the heart of the rule of law lies the principle of legality, the requirement that public authority be exercised in accordance with pre-existing, clear, and foreseeable legal norms.[58] For AI governance, legality implies that the use of algorithmic systems in anti-corruption institutions be explicitly authorized by law, delineating their purpose, scope, and limits.[59] The Framework Convention codifies this principle by obliging states parties to put in place measures that ensure AI systems comply with rule-of-law standards.[60]

In practice, legality serves a dual function: it safeguards individuals from arbitrary technological decisions and ensures that public authorities remain legally accountable for their deployment and use.[61] The Venice Commission clarifies that legality entails both express statutory authorization and the availability of judicial review, requiring that individuals affected by administrative or judicial decisions have effective access to remedies.[62] This principle is particularly crucial where AI is used to detect or prioritize corruption investigations, since automated tools may influence prosecutorial discretion, a domain traditionally governed by law and human judgment.

The principle of legal certainty, closely related to legality, further requires that laws regulating AI be accessible, clear, concise, and foreseeable in their application. Besides laws, anti-corruption agencies should also adopt clear procedural rules concerning data collection, model training, and system outputs.

4.2. Legitimate purpose, necessity, and proportionality tests

The legitimate purpose ensures that AI is used only to pursue objectives consistent with democratic values and human rights.[63] Within anti-corruption frameworks, legitimate purposes may include fraud detection, mapping of conflicts of interest, risk analysis, and administrative transparency. Uses of AI for politically motivated targeting, surveillance, or reputational manipulation fall outside this scope and contravene the rule of law.[64]

Necessity and proportionality, long-standing principles in both international law and administrative law, require that any interference with fundamental rights and freedoms "be necessary in a democratic society and proportionate to the legitimate aim pursued."[65] Applied to AI, this means that automated systems should only be used where manual or less intrusive methods are insufficient to achieve anti-corruption objectives. Moreover, the intensity of algorithmic intervention should correspond to the level of risk, reflecting the "risk-based approach" enshrined in Article 16 of the CoE's Framework Convention on AI.[66]

Proportionality also governs data processing. On this basis, AI systems can only collect and process data that are strictly necessary for their specified, explicit, and legitimate anti-corruption purposes.[67] The General Data Protection Regulation of the EU offers a useful contextual benchmark, requiring that data be "adequate, relevant and limited to what is necessary."[68] This standard, when applied to AI-driven investigations or audits, helps balance efficiency with the protection of privacy and due process.

4.3. Transparency and explainability

Transparency and explainability constitute core procedural pillars of lawful and accountable AI governance. The CoE's Framework Convention on AI, same as the EU's AI Act,

58 Venice Comm'n, Rule of Law Checklist, *supra* note 5.

59 *Id.*

60 Framework Convention on AI, *supra* note 4, art. 1.

61 Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1147 (2017).

62 Venice Comm'n, Rule of Law Checklist, *supra* note 5.

63 Framework Convention on AI, *supra* note 4, pmbl.

64 Crawford, *supra* note 21.

65 See, e.g., European Convention on Human Rights, arts. 8(2), 9(2), 10(2) & 11(2) and the practice of the European Court of Human Rights. For a number of earlier, classic cases in the Court's jurisprudence, see: *Handyside v. United Kingdom*, 24 Eur. Ct. H.R. (ser. A) (1976); *The Sunday Times v. United Kingdom (No. 1)*, 30 Eur. Ct. H.R. (ser. A) (1979). See also: Nina Peršak, *Positive Obligations in View of the Principle of Criminal Law as a Last Resort*, *in* COERCIVE HUMAN RIGHTS 187 (Laurens Lavrysen & Natasa Mavronicola eds., Hart Publ'g 2020).

66 Framework Convention on AI, *supra* note 4, art. 16.

67 Regulation 2016/679, 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 and on the Free Movement of Such Data (General Data Protection Regulation), art. 5(1)(b), 2016 O.J. (L 119) 1 [hereinafter GDPR].

68 *Id.*

enshrine transparency as a legal obligation, requiring states to make sure that individuals are informed or aware of the circumstances when they are interacting with AI systems,[69] and have access to understandable explanations of AI-assisted decisions.[70] In anti-corruption institutions, transparency operates at two principal levels: (1) institutional transparency, concerning public disclosure of AI deployment, and (2) individual transparency, concerning the ability of affected persons to understand algorithmic outputs that impact them.

Institutional transparency requires anti-corruption agencies to disclose, at minimum, the general purpose, design, and limitations of the AI systems they utilize. Public registers of algorithmic tools, similar to the EU's database designed for high-risk AI systems under the AI Act, could enhance accountability and public trust.[71] Individual transparency, by contrast, concerns due process, implicating that where AI influences a decision about a person, that person must be able to understand the reasoning, contest the outcome, and seek redress.[72]

Explainability, often referred to as explainable AI (XAI),[73] is essential to align technical processes with legal accountability.[74] Anti-corruption contexts pose unique challenges because predictive analytics or anomaly detection systems may rely on opaque statistical inferences. The principle of transparency thus requires that such systems be interpretable by humans, particularly decision-makers and judicial authorities.

4.4. Accountability and human oversight

Accountability ensures that responsibility for AI-supported decisions remains identifiable or traceable.[75] Effective accountability depends on traceability, meaning that AI systems ought to be structured in ways that decisions, data inputs, and model parameters can be audited ex post.[76] In anti-corruption agencies, traceability is vital to prevent the use of opaque algorithms to justify politically expedient decisions or conceal improper influence.

The CoE's Framework Convention explicitly mandates oversight across the AI lifecycle, requiring states to establish oversight frameworks proportioned to particular contexts and risk levels.[77] It elevates human oversight from a policy recommendation to a binding legal obligation, emphasizing that states must ensure human control consistent with risks and contexts.[78] In the anti-corruption context, the specialized anti-corruption agencies or bodies should establish AI ethics committees or independent supervisory bodies tasked with auditing AI systems, approving high-risk deployments, and ensuring compliance with human rights standards.

4.5. Equality and non-discrimination

One of the core elements of the rule of law is equality before the law and prohibition of discrimination.[79] AI systems trained on biased data or developed without diversity safeguards risk violating this foundational norm,[80] particularly since "automating fairness … may be impossible."[81]

The Framework Convention obliges states to establish and uphold measures ensuring that AI systems comply with the principles of equality and non-discrimination under international and national law.[82] As consistently held by the European Court of Human Rights, discrimination includes both direct and indirect forms.[83]

In anti-corruption settings, algorithmic bias may manifest through data selection, model training, or interpretative assumptions. For instance, systems that predict corruption risks based on socioeconomic variables may inadvertently stigmatize certain regions or professions. To prevent such scenarios or results, AI systems should undergo regular bias audits, incorporate datasets, and ensure inclusive design processes involving multidisciplinary expertise.

The principle of non-discrimination extends beyond model fairness to institutional practice. AI systems, being

69 Framework Convention on AI, *supra* note 4, art. 15(2).
70 AI Act, *supra* note 50, ¶ 171.
71 *Id.* art. 71.
72 Framework Convention on AI, *supra* note 4, art. 14(2).
73 David Gunnning & David W. Aha, *DARPA's Explainable Artificial Intelligence (XAI) Program*, AI MAG. 44, 44 (2019).
74 Selbst & Barocas, *supra* note 19.
75 OECD, *Council Recommendation on Artificial Intelligence*, *supra* note 46, ¶ 1.5.
76 *Id.*
77 Framework Convention on AI, *supra* note 4, art. 8.
78 *Id.* art. 8.
79 Venice Comm'n, Rule of Law Checklist, *supra* note 5.
80 Wachter, Mittelstadt & Russell, *supra* note 18.
81 *Id.*
82 *Id.* art. 10
83 See, e.g., D.H. & Others v. Czech Republic, 2007-IV Eur. Ct. H.R. 241.

"expressions of power … from wider economic and political forces," must not be used selectively to target political opponents or shield favored groups.[84] The Rule of Law Checklist frames equality as a component of the prohibition of arbitrariness or any unjustified unequal treatment under the law, reinforcing that impartiality in governance and in the use or operation of AI is integral to the rule of law.[85]

4.6. Independence and integrity of anti-corruption bodies

The independence of specialized anti-corruption bodies is a central tenet of both UNCAC and CoE/Venice Commission standards.[86] States are required to ensure that such bodies possess the requisite independence to perform "their functions effectively and free from any undue influence."[87] Apart from specialized staff and resources, independence ensures that investigations and prosecutions are protected from political interference.[88] When AI systems are introduced into these institutions, they must not undermine this independence by creating new dependencies, be it technological, financial, or informational.

Vendor-provided AI solutions, for example, may give private actors undue influence over investigative priorities or data management.[89] Similarly, algorithmic tools that rely on politically designed datasets risk compromising institutional integrity. To safeguard independence, anti-corruption agencies or other specialized bodies should retain control over algorithmic systems, with transparent procurement processes.

4.7. Data protection and privacy

Data protection is both a human right and a functional precondition for legitimate use of AI systems. Both the Framework Convention and the European Convention on Human Rights (ECHR) recognize privacy as integral to democratic governance.[90]

AI systems used in anti-corruption efforts will inevitably process large datasets, including personal and sensitive information. Given this, they must be compliant with data protection standards such as the ones enshrined in the EU's General Data Protection Regulation (GDPR) and the CoE's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.[91]

Data minimization, purpose limitation, and security safeguards are key obligations that should govern AI systems in the anti-corruption context.[92] In any event, they should at any time be legally, fairly and transparently processed in connection to the data subject.[93] Moreover, AI systems used for predictive or risk-scoring purposes must avoid profiling that infringes individual rights and freedoms or presumes guilt without due process.[94] Therefore, in order to balance transparency and confidentiality, anti-corruption agencies may implement tiered access models, where sensitive data are pseudonymized or aggregated before analysis. Privacy protection in AI systems for anti-corruption is not merely a matter of legal compliance, it is also essential for maintaining public legitimacy and fostering trust in public institutions. Citizens must trust that AI tools will not become instruments of mass surveillance.

4.8. Competence and capacity building

Capacity constraints constitute a decisive variable in the responsible deployment of AI in anti-corruption contexts. In low-resource jurisdictions, limited technical expertise, budgetary constraints, and dependence on external vendors may heighten risks to institutional independence and accountability.

A principled AI framework must be sensitive to asymmetries in institutional capacity, emphasizing scalable safeguards, shared oversight mechanisms, and international technical assistance rather than one-size-fits-all compliance models. One possible modality to expand AI's potential and make it more inclusive is through broad and diverse international partnerships. For instance, at France's February 2025 AI Action Summit, it was announced that ten countries[95] have launched a Public Interest AI Platform and Incubator to support, amplify, and decrease fragmentation between

84 Crawford, *supra* note 21, at 211.
85 Venice Comm'n, Rule of Law Checklist, *supra* note 5.
86 United Nations Convention against Corruption, *supra* note 16, art. 6.
87 *Id.*
88 *Id.*
89 Engstrom et al., *supra* note 20.
90 European Convention on Human Rights, *supra* note 65, art. 8; Framework Convention on AI, *supra* note 4, art. 14.
91 Council of Eur., Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, Eur. T.S. No. 108.
92 GDPR, *supra* note 67, art. 5(1).
93 *Id.*
94 *Id.* art. 22.
95 These countries are: Chile, Finland, France, Germany, India, Kenya, Morocco, Nigeria, Slovenia, and Switzerland.

existing public and private initiatives on public interest AI, and address digital divides.[96] This Public Interest AI Platform and Incubator will support digital public goods, technical assistance and capacity building to foster a trustworthy ecosystem for public interest AI.[97]

On competence, the EU has developed a comprehensive competency framework to guide public servants in adopting and managing AI effectively,[98] which can be adapted or replicated elsewhere. The framework classifies competencies into three main dimensions: technical competencies, managerial competencies, and policy, legal and ethical competencies.[99] The framework thus highlights the need for a structured approach to AI skills development, emphasizing technical, managerial and legal and policy-oriented competencies. Aligning with these principles, Canada's AI Strategy for the Federal Public Service 2025-2027 focuses on four priority areas, which is central AI capacity (a central hub to support project implementation and share knowledge); policy, legislation, and governance; talent and training; and engagement, accountability, transparency, and value to Canadians.[100]

4.9. Independent evaluation and effective remedies

The rule of law requires mechanisms of independent evaluation and effective remedy. The Framework Convention on AI obliges states to guarantee that individuals affected by AI systems have access to remedies that are both effective and readily accessible.[101] In such circumstances, states are obliged to provide for persons affected effective remedies against public authorities.[102] This may involve administrative appeals, judicial review, or independent audit mechanisms capable of ordering correction or cessation of unlawful AI practices.

Institutionally, states should adopt legislative and other measures to establish independent review mechanisms (judicial, parliamentary or ombudsperson-based) to oversee the legality and proportionality of AI operations in the anti-corruption context. As a measure of due diligence, algorithmic impact assessments should be mandated to evaluate legality, human rights compliance, and effectiveness.[103] Such assessments, designed along the lines of environmental impact evaluations, would enable continuous learning and adaptation, strengthening institutional accountability.

Finally, all these principles, taken together, form an integrated normative framework that enables and ensures the use of AI in anti-corruption governance remains within constitutional limits and in compliance with international legal obligations, primarily those concerning human rights.

Compliance with the proposed framework may be assessed through qualitative and quantitative indicators, including the existence of legal authorization, frequency of human override, audit findings, bias metrics, and accessibility of remedies. A typical remedy pathway would involve notification of AI use, access to explanation, administrative appeal within the anti-corruption body, and judicial or independent review of ordering correction or termination of unlawful practices.

4.10. Illustrative case vignettes: Opportunities, risks, and safeguards

4.10.1. AI-based procurement risk scoring

Several States have introduced AI-driven tools to detect corruption risks in public procurement by analyzing bid patterns, pricing anomalies, and supplier networks.[104] Such systems can enhance early detection and reduce discretion at the administrative level. However, without safeguards, risk-scoring tools may disproportionately flag small or regional suppliers, reinforcing structural inequalities.

Applying the principles outlined in this article, such systems should be legally authorized, limited to risk prioritization (not automatic exclusion), and subject to human validation. Transparency requirements demand public disclosure of the system's purpose and general logic, while affected bidders must have access to explanations and remedies. Independent audits should assess whether risk indicators correlate with protected characteristics or geographic bias.

96 *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet*, Élysée (Feb. 11, 2025), https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet.

97 *Id.*

98 R. Medaglia, P. Mikalef & L. Tangi, *Competences and Governance for Artificial Intelligence in the Public Sector*, Publ'ns Off. of the Eur. Union (2024), https://data.europa.eu/doi/10.2760/7895569.

99 *Id.*

100 *AI Strategy for the Federal Public Service 2025-2027*, Gov't of Can., https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/gc-ai-strategy-overview.html (last visited 6 July 2026).

101 Framework Convention on AI, *supra* note 4, art. 14.

102 *Id.* art. 14(2)(c).

103 Dillon Reisman, Jason Schultz, Kate Crawford & Meredith Whittaker, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* (2018), https://ainowinstitute.org/wp-content/uploads/2023/04/aiareport2018.pdf.

104 See: OECD, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions, supra* note 7, at 204-12 (AI in public procurement).

This example shows how AI can support UNCAC-compliant procurement integrity while remaining subordinate to human judgment, due process, and public accountability.

4.10.2. AI-based risk scoring in asset declarations: The Armenian experience

As part of its anti-corruption reforms, Armenia established the Corruption Prevention Commission (CPC) in 2019 and entrusted it with overseeing the electronic register of asset and interest declarations submitted by public officials.[105] Faced with a high volume of declarations and limited institutional capacity, the CPC initially struggled to conduct meaningful verification, particularly given that early submission were not machine-readable. To address these constraints, the CPC developed a data platform that structured declaration data and enabled interoperability with databases held by other public authorities.[106]

Building on this infrastructure, the CPC introduced an automated verification system to conduct preliminary screening of declarations and identify potential red flags.[107] The system is currently being complemented by a pilot AI component designed to learn from prior verification outcomes and detect emerging patterns of corrupt behavior. Consistent with UNCAC obligations concerning preventive anti-corruption bodies and integrity systems[108] and in line with rule of law principles elaborated in this article, the system is confined to risk prioritization and does not generate automated findings of illegality or sanctions; all follow-up decisions remain subject to human review and institutional accountability.

The development process illustrates the importance of legality, proportionality, and participation in governing AI use in anti-corruption institutions. Stakeholder consultations with domestic and international private-sector actors informed both the technical design of the system and the sequencing of reforms, leading the CDC to prioritize data quality and governance before deploying machine-learning techniques.[109] At the same time, ethical and legal tensions emerged between transparency and privacy. While Armenian law, which is aligned with the GDPR, permits publication of asset declarations in the public interest, the CPD concluded that full disclosure of the verification algorithm could compromise personal data protection and the integrity of oversight processes.[110]

To mitigate these risks, the CPC has emphasized procedural transparency by explaining the purpose, scope, and general logic of the system, without disclosing sensitive technical details. Individuals whose declarations are flagged are entitled to notice and an opportunity to respond, preserving due process and trust. This case demonstrates how AI can enhance preventive anti-corruption enforcement when embedded within clear legal mandates, robust data governance, and safeguards that ensure human oversight, privacy protection, and public accountability.

To comply with the proposed framework in this context, any AI system should operate under lawfully enacted mandates, require human confirmation of high-risk flags before further action (making sure that algorithms make no ethical decisions), and ensure that individuals flagged by the system are informed and given an explanation with access to redress.

5. ELEMENTS OF AN INTERNATIONAL DECLARATION ON THE USE OF AI IN ANTI-CORRUPTION

The preceding analysis shows that existing international frameworks, most notably the Framework Convention on AI, the Rule of Law Checklist, and related OECD, UNODC, and EU's AI Act standards, together provide a coherent foundation for responsible AI governance. Despite these convergences, however, there is no dedicated normative instrument articulating how such principles should specifically govern the use of AI within anti-corruption laws, policies, and institutions.

This normative gap is both practical and conceptual. On the one hand, anti-corruption agencies, integrity commissions, and judicial bodies increasingly employ AI for case prioritization, data mining, risk assessment, and public procurement monitoring. Yet these systems operate in a regulatory vacuum, often without explicit legal authorization, oversight, or public accountability. On the other hand, theoretical work on the intersection of AI and the rule of law has primarily focused on general governance contexts, such as public administration, policing, or integrity of electoral processes, rather than the specialized mandates and sensitivities of anti-corruption enforcement.

Given the speed of development and deployment of AI and complexity and potential for abuse, a guiding framework becomes a central imperative. In this connection,

105 OECD, *Generative AI for Anti-Corruption and Integrity in Government: Taking Stock of Promise, Perils and Practice* 37 (2024).
106 *Id.*
107 *Id.*
108 United Nations Convention against Corruption, *supra* note 16, arts. 6, 36.
109 OECD, *Generative AI for Anti-Corruption and Integrity in Government: Taking Stock of Promise, Perils and Practice, supra* note 105, at 37.
110 *Id.*

there is clear value in elaborating a set of core principles in a coherent declaration or statement setting out common principles to guide States, institutions, and scholarly debate. Declarations have historically served as norm-generative mechanisms, articulating standards that later crystallize into either custom or treaty obligations, or both.[111] The Universal Declaration of Human Rights of 1948,[112] for example, laid the groundwork for legally binding conventions, namely the International Covenant on Civil and Political Rights and International Covenant on Economic, Social and Cultural Rights. Similarly, the proposed Declaration of Core Principles on the Use of AI in Anti-Corruption Laws, Policies, and Institutions could serve as a soft law bridge between the technical design of AI and the constitutional values of legality, equality, and institutional independence.

This Declaration would be conceived not as a parallel instrument to existing frameworks, but rather as a complementary articulation of how general applicable principles to AI systems, such as legality, transparency, or accountability should be operationalized within the specific institutional ecosystem of anti-corruption governance. It would underscore the continuity between AI ethics and the rule of law, positioning AI not as a technological disruption but as a governance tool that must governed by, or remain subordinate, to democratic principles and processes and human oversight.[113]

On the basis of the normative framework as elaborated in the preceding section, the Declaration identifies ten clusters of normative principles derived from existing international standards and practice, including scholarly work:

1. Legality and Legal Certainty: Reaffirming that AI must operate within clear and foreseeable legal norms, enacted through democratic procedures.[114]
2. Legitimate Purpose, Necessity, and Proportionality: Ensuring that AI use is justified, proportionate, and directed toward legitimate anti-corruption objectives.[115]
3. Transparency and Explainability: Mandating public disclosure of AI use and explainability of AI-supported decisions.[116]
4. Accountability and Human Oversight: Preserving human responsibility and legal liability across the AI lifecycle.[117]
5. Equality and Non-Discrimination: Preventing algorithmic bias and ensuring fairness and non-discrimination in all AI-enabled processes.[118]
6. Independence and Integrity: Protecting anti-corruption bodies from political interference through AI systems.[119]
7. Data Protection and Privacy: Requiring observance of international data protection and privacy standards.[120]
8. Competence and Capacity Building: Promoting institutional readiness and AI literacy among public officials and citizens.[121]
9. Independent Evaluation and Effective Remedies: Guaranteeing external audits and effective remedies for affected individuals.[122]
10. International Cooperation and Best Practices: Encouraging harmonized, rights-based approaches and exchange of expertise.[123]

By embracing and consolidating these principles, the Declaration seeks to align technical innovation with institutional integrity, thereby contributing to a rule of law-centered AI governance model. It offers practical guidance for states drafting national AI policies, anti-corruption agencies adopting digital tools, and international organizations and civil society developing evaluation criteria.

The full text of a Draft Declaration of Core Principles on the Use of AI in Anti-Corruption Laws, Policies, and Institutions is annexed to this article. It is presented as a contribution to ongoing international discussions about ethical, legal, and institutional standards for the responsible deployment and use of AI in anti-corruption contexts.

111 See, e.g., Louis B. Sohn, *The Development of the Charter of the United Nations: A Commentary,* 39 YALE L.J. 1177 (1951).
112 See: Universal Declaration of Human Rights, G.A. Res. 217A (III), U.N. Doc. A/810 (Dec. 10, 1948).
113 Framework Convention on AI, *supra* note 4.
114 Venice Comm'n, Rule of Law Checklist, *supra* note 5.
115 Framework Convention on AI, *supra* note 4; European Convention on Human Rights, *supra* note 65.
116 Framework Convention on AI, *supra* note 4, art. 8; AI Act, *supra* note 50, ¶ 171.
117 Framework Convention on AI, *supra* note 4, arts. 8-9.
118 *Id.* art. 10.
119 *Id.* art. 5; United Nations Convention against Corruption, *supra* note 16, art. 6.
120 Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, *supra* note 91; GDPR, *supra* note 67.
121 Framework Convention on AI, *supra* note 4, art. 20.
122 *Id.* art. 14.
123 *Id.* art. 25; OECD, *Council Recommendation on Artificial Intelligence*, *supra* note 46, ¶ 2.5.

6. CONCLUSION

AI is transforming the operational and normative landscape of public governance, particularly in the fight against corruption. Yet, as this article has shown, integrating AI into anti-corruption efforts raises fundamental human rights and rule-of-law concerns. The pace or speed of development and evolution of AI systems cannot be permitted to outpace the legal and ethical frameworks that safeguard democratic integrity. Without principled governance, AI's deployment in this field risks not only inefficiency but also institutional distortion and abuse.

The CoE's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law marks a monumental moment and turning point in global AI governance. It offers a legally binding articulation of principles that have, until now, remained largely aspirational, such as transparency, accountability, human oversight, and proportionality. By situating AI regulation within the normative architecture of human rights and democratic governance, the Framework Convention underscores that the technological revolution must remain subject to rule of law and public accountability. The Rule of Law Checklist of the Venice Commission further complements this approach by grounding AI governance within well-established rule of law parameters, as identified and discussed in this article. Such rule of law principles are operational benchmarks that help ensure that AI systems used in anti-corruption work reinforce, rather than erode, the institutional independence and procedural fairness that are indispensable for effective governance.

Beyond Europe, global and regional organizations, such as the OECD, UNODC, and the EU, have contributed to a growing corpus of soft and hard law instruments that codify ethical and procedural safeguards for AI. Together, these frameworks signify an emerging convergence around a human-centered model of AI regulation, emphasizing transparency, explainability, and accountability. It is in this context that this article advances a normative framework tailored to anti-corruption efforts. It argues that the responsible deployment and operation of AI should be governed by a set of foundational legal principles.

In illustrating the application of governing principles through selected institutional scenarios, this article seeks to bridge the gap between abstract norm-setting and operational governance. Embedded in UNCAC and complementary international standards, the proposed framework is intended to remain adaptable across diverse legal systems and institutional settings.

The Draft Declaration of Core Principles on the Use of Artificial Intelligence in Anti-Corruption Laws, Policies, and Institutions, annexed to this article, is proposed as a tool for dialogue among states, policymakers, and scholars. It seeks to translate complex normative principles into an actionable set of commitments. The ultimate goal is to embed a rule-of-law culture into the design, deployment and use of AI tools before harm occurs.

In sum, AI has the potential to become an ally, rather than a threat, to integrity and transparency in governance. Notwithstanding its potential, AI governance in anti-corruption contexts entails structural tensions. Automation may enhance efficiency while simultaneously narrowing discretion; transparency may conflict with investigative confidentiality; and reliance on predictive tools may subtly reshape enforcement priorities. These tensions underscore that AI cannot be a substitute for institutional integrity, but only a contingent tool whose legitimacy and value depends on human judgment, oversight, and democratic accountability.

Ultimately, realizing AI's potential responsibly demands a deliberate and principled approach, one that harmonizes technological innovation with the enduring values of human rights, democracy, and the rule of law. These values, once transformed into a consolidated normative foundation, represent a step toward positioning AI as a guardian, not a challenger, to the rule of law in global anti-corruption efforts and frameworks.

Annex 1: Draft Declaration of Core Principles on the Use of Artificial Intelligence in Anti-Corruption Laws, Policies, and Institutions

Rationale

The accountable and responsible use of artificial intelligence (AI) offers powerful tools in the fight against corruption, but it must be guided by democratic values, legal safeguards, and institutional integrity. These principles aim to ensure that anti-corruption AI systems promote justice, fairness, and trust in public institutions.

Preamble

Recognizing the growing role of AI systems in the fight against corruption;
Affirming the necessity that the deployment of AI systems in anti-corruption frameworks respects the rule of law, human rights, and democratic standards;
Noting the United Nations Convention against Corruption, the Council of Europe's Framework Convention on Artificial

Intelligence and Human Rights, Democracy and the Rule of Law, and the Venice Commission's standards, including in particular its Rule of Law Checklist;
We (the undersigned), hereby declare the following core principles governing the use of AI in anti-corruption efforts, notably within specialized anti-corruption agencies and courts:

I. Legality and Legal Certainty

1. The design, deployment and operation of AI systems in anti-corruption contexts shall be grounded in clear, accessible, and foreseeable legal norms.

2. Laws regulating AI in anti-corruption institutions must be adopted through democratic processes and subject to human oversight and public scrutiny.

II. Legitimate Purpose, Necessity and Proportionality

3. AI systems shall only be used to pursue clearly defined and legitimate anti-corruption objectives, such as detection, prevention, case prioritization, or risk analysis.

4. The use of AI must be necessary and proportionate to the goal pursued.

III. Transparency and Explainability

5. The deployment of AI in anti-corruption institutions must be publicly disclosed, including the general purposes, capabilities, and limitations of such systems.

6. AI-driven decisions that affect individuals or legal proceedings must be explainable in terms understandable by humans, especially parties to a proceeding.

IV. Accountability and Human Oversight

7. Decisions with legal or procedural consequences, particularly in judicial or prosecutorial processes, shall not be taken solely by AI. Human officials must remain responsible and retain final decision-making authority.

8. The data sources, processes, and decisions of AI systems must be documented and auditable by independent bodies.

V. Equality and Non-Discrimination

9. AI systems ought to be designed and tested constantly to prevent direct or indirect discrimination based on race, gender, ethnicity, political affiliation, or any other prohibited ground.

10. The development of AI tools should include safeguards to ensure accessibility and fairness, especially for vulnerable or marginalized groups.

VI. Independence and Integrity of Anti-Corruption Bodies

11. The use of AI cannot undermine the operational, investigative, or adjudicative independence of specialized anti-corruption bodies or courts.

12. AI tools must not be used to target or shield individuals or groups for political purposes. Use must be impartial and resistant to political manipulation.

VII. Data Protection and Privacy

13. AI systems shall process data in full compliance with applicable data protection standards and must respect privacy rights of individuals.

14. Only data strictly necessary for the anti-corruption purpose shall be processed by AI systems.

VIII. Competence and Capacity Building

15. Staff in anti-corruption bodies must be adequately trained in the use, oversight, and limitations of AI tools.

16. Public institutions shall promote awareness of how AI is used in anti-corruption efforts, to foster trust and democratic legitimacy.

IX. Independent Evaluation and Effective Remedies

17. Independent bodies should be empowered to review AI systems used in anti-corruption contexts for legality, ethics, and effectiveness.

18. Individuals impacted by AI-assisted decisions should be able to seek effective remedies, including judicial and/or constitutional review.

X. International Cooperation and Best Practices

19. States and agencies shall cooperate to promote interoperable, rights-based standards for AI in anti-corruption work.

20. Institutions should share best practices, risk assessments, and evaluation reports to strengthen responsible AI governance globally.

AI Use Disclosure

The author declares full intellectual authorship of this article. ChatGPT was used only for language proofreading, with all edits reviewed by the author.

Disclosure of interest

The author has no competing interests to declare.