TRANSPARENCY INTERNATIONAL
the global coalition against corruption

ADDRESSING CORRUPT USES OF ARTIFICIAL INTELLIGENCE

Transparency International is a global movement with one vision: a world in which government, business, civil society and the daily lives of people are free of corruption. With more than 100 chapters worldwide and an international secretariat in Berlin, we are leading the fight against corruption to turn this vision into reality.

www.transparency.org

Working paper

Addressing Corrupt Uses of Artificial Intelligence

Cover: Transparency International with assets from Freepik

ISBN: 978-3-96076-275-1

ADDRESSING CORRUPT USES OF ARTIFICIAL INTELLIGENCE

Recent advancements in artificial intelligence (AI) are transforming numerous sectors at an unprecedented speed. Automated decision making in government, for example, has the potential to enhance efficiency and consistency in public services, such as tax administration, social security systems and regulatory compliance. By leveraging AI, governments can process large volumes of data quickly, reduce human error and provide timely responses to citizens. AI can also contribute to preventing corruption in service delivery by reducing the risk of human intervention or abuse, and by institutionalising compliance rules.

However, while both the development and deployment of AI continue to grow largely unimpeded, lawmakers are slow to regulate its use. Many reports show a drastic implementation lag, as AI regulations roll out more slowly than technological advances.[1][2] Transparency International's *Global Strategy 2021-2030* recognises that developments in artificial intelligence, big data, cryptocurrencies and social media are set to further change the landscape of political and corporate influence, offering not only new tools for transparency and accountability, but new forms, vehicles and opportunities for corruption.[3] The global strategy also commits us to work on ensuring that "automation, artificial intelligence and new technologies are accountable and used to secure the common good – not as new mechanisms for exclusion, deprivation of public entitlements or arbitrary decision making".[4]

Transparency International is concerned specifically with the corrupt use of AI, defined as the intentional abuse of AI systems for the private benefit of entrusted powerholders. This is distinct from issues arising from unintended consequences and outcomes of the adoption of AI systems. In particular, this working paper covers the corrupt use of AI in the public sector, which can take many forms. For example, in resource allocation, manipulated AI systems might disproportionately direct funds to particular regions or political constituencies. In policy implementation, AI-driven models could be skewed to support predetermined outcomes, undermining fair and evidence-based policymaking. In public procurement, AI systems could be intentionally programmed to unfairly prioritise certain vendors. Generative AI can be abused by corrupt actors to fuel disinformation campaigns that undermine political integrity. AI can also be used to enable corrupt actors to amplify disinformation campaigns, manipulate public discourse, and enhance censorship and surveillance, thereby undermining civil and political rights. These abuses of AI intensify existing threats to transparency and fairness in governance, particularly in the public sector.[5]

Additionally, while many governmental, non-governmental and private-sector organisations are active in the AI integrity and ethics space, there is currently no deliberate focus on, or call for, measures that ensure AI is not used for corrupt outcomes or to facilitate corrupt acts. At the policy level, the current discourse generally overlooks the intentional misuse of AI. This working paper therefore sheds light on how corruption relates to AI integrity, by defining the corrupt use of AI, providing overviews of the main manifestations of such use, and making recommendations on how to address these challenges.

TRANSPARENCY INTERNATIONAL'S DEFINITION OF CORRUPT USE OF AI

This working paper is a first and foundational step in Transparency International's work to prevent and counter corrupt uses of AI. In this working paper, "AI systems" refer to any elements that are part of the commissioning, deployment and evaluation of AI tools. These range from the procurement process of the AI system, the code and algorithms used, the training data – the initial dataset used to train machine learning algorithms in an AI system, and the impact evaluations and assessments conducted pre- and post-deployment. This working paper primarily focuses on the corrupt use of AI within the public sector, whether developed internally or sourced from private companies.

Transparency International defines corruption as the abuse of entrusted power for private gain. This definition has stood the test of time, as it has allowed for a broad array of manifestations of corruption, including in the public and private sectors.[6] In addition to this definition of corruption, Transparency International's 2021-2030 Strategy introduces the concept of holding power to account for the common good. The "common good" refers to the United Nations Sustainable Development Goals as a universal set of socially optimal targets, with powerholders ranging from the more traditional, such as public and elected officials, to new actors like large tech companies and media platforms.[7]

Transparency International's definition of "the corrupt use of AI" draws from concepts in the definition of corruption and our global strategy goal of holding power to account for the common good. We define the corrupt use of AI as the intentional abuse of AI systems for the private benefit of entrusted powerholders. Corrupt uses of AI include both (i) the intentional manipulation of AI systems themselves for corrupt outcomes, and (ii) the use of AI tools in the commission of corrupt or corruption-related acts. The term "abuse" implies some degree of purpose, as this working paper emphasises that intentionality matters when considering an act of corruption, especially since the outcomes of AI systems can be unpredictable and at times beyond their users' control. Therefore, public policy failures that are a result of unintended bias or other forms of human error would not fall under the definition of the corrupt use of AI. For example, a case in which a contracting authority intentionally selected training data to ensure that procurement tenders would go to certain companies would constitute corrupt use of AI. However, if the training data available happens to favour certain companies and the authority has not intentionally selected it for this purpose, this would not be considered a case of corruption.

FEATURES THAT MAKE AI PRONE TO INTENTIONAL ABUSE

Opacity in AI systems' decision making is one of the key features that enable intentional abuse of such systems. The complexity of "black box" algorithms, whose outputs often defy simple explanations, means that even programmers struggle to trace or comprehend the reasoning behind the decisions. In addition, the datasets on which an algorithm is trained often remain hidden from public scrutiny.[8] AI's ability to process large datasets and make highly accurate predictions often gives an illusion of neutrality, as algorithms appear to eliminate the human factor in decision making.[9] This issue is compounded in some legal systems, where evidence produced by computers may be assumed to be reliable unless proven otherwise.[10] This perceived neutrality, paired with opacity resulting from complex algorithms, poses a risk, as decisions made by a corrupt system may appear legitimate, leading to potential abuse in a range of areas, including criminal justice, health care and public services.[11]

AI can also concentrate power in the hands of a few, amplifying risks of abuse through opaque decision making and discriminatory algorithms. As in other technological domains, the high degree of specialisation, expertise and resources required to develop AI systems results in the dominance of a small number of powerful tech companies. Their specialist knowledge and control over AI systems make them powerful gatekeepers, who require adequate oversight. In regimes where power is highly concentrated, AI-driven surveillance tools have been extensively used to suppress dissent and erode political accountability.[12][13]

Low levels of transparency and oversight in the deployment and procurement of AI systems also introduces risks. A lack of information about the uses of AI within the public sector – including the types of decisions, services or policies informed by AI-driven decision making – prevents proper scrutiny by the public. The procurement of such

systems is also typically carried out with little to no opportunities for oversight. Without clear procurement guidelines and oversight, companies could exploit these integrity gaps to exert undue influence over public officials, creating risks of manipulated decision making.[14]

At an operational level, AI can enable corrupt activities by automating tasks that would usually require a larger, more exposed network. For instance, AI can streamline the establishment of companies, or the creation of fraudulent documents or bids in procurement processes, enhancing the efficiency and secrecy of corrupt acts.[15] Generative AI and large language models (advanced artificial intelligence systems designed to understand and generate human language) in particular present new risks – for example, with threat actors creating fraudulent documentation and AI-generated fake identities to bypass know-your-customer obligations.[16]

MANIFESTATIONS OF CORRUPT USE OF AI

INTENTIONAL MANIPULATION OF AUTOMATED DECISION MAKING

Intentional manipulation can stem from various human factors, including the inputs and the diversity of training data used in algorithmic processes. There are at least three key areas where intentional manipulation can occur in algorithmic decision-making systems. Firstly, corrupt actors can manipulate data selection by choosing data that is not representative – for example, over- or under-recording specific groups, leading to skewed outputs. Secondly, such actors can manipulate algorithmic design through deliberate choices that favour certain groups over others. Lastly, individuals interpreting or using algorithmic outputs can introduce intentional bias by applying their own conscious biases to the final decision – so-called human oversight manipulation.[17,18] For example, dishonest data scientists designing AI systems to predict patient survival and make treatment suggestions could use selective training data to favour themselves, their peers or those who can afford treatment (data selection manipulation). Government officials contracting the design of such an AI system could insist on algorithmic tweaks that favour companies of party donors (design manipulation). The doctors overseeing the use of the system could, in turn, apply additional filters that favour patients with private health insurance (human oversight manipulation). Or they might bias the AI to discriminate based on demographics like age or race, ensuring preferential treatment for their group in emergencies.[19]

A particularly harmful form of corrupt use of AI is "algorithmic capture", which refers to AI systems being manipulated to systematically favour a specific group – for example, by manipulating the code of algorithms used in electronic procurement or fraud detection to steer lucrative contracts to cronies or to conceal wrongdoing.[20]

USE OF AI FOR LAUNDERING CORRUPTION-RELATED FUNDS

Corruption-related illicit financial flows have a profound negative impact on societies – for example, by reducing the ability to fund sustainable and inclusive development, undermining the rule of law, or allowing corrupt elites to use the global financial system to launder the proceeds of their wrongdoing. Kleptocrats and corrupt public officials rely on money laundering techniques to conceal their corrupt acts, avoid public scrutiny and use their ill-gotten gains for malign influence or luxury consumption. With increasing uptake of AI systems across a broad spectrum of financial services,[21] these systems become vulnerable to abuse by corrupt actors to launder the proceeds of corruption. In parallel, AI systems can be used to facilitate money laundering by corrupt actors and their enablers.

For example, some governments have expressed concerns about the use of AI to further fraud, which also poses a threat to combatting corruption-related money laundering. Governments have specifically highlighted the use of generative AI and large language models to generate fraudulent documentation or create synthetic identities to bypass compliance procedures.[22] This misuse of AI systems can also facilitate laundering of the proceeds of corruption. Corrupt actors laundering large amounts of illicit gains often employ large, complex schemes involving numerous shell companies, trusts and various assets. Fraudulent documentation is frequently required

to justify transactions to an obliged entity – an entity or person subject to anti-money laundering obligations – in such a scheme.[23] Generative AI systems could therefore pose a substantial risk of corruption, due to their ability to quickly produce fraudulent documents, which can help bypass due diligence checks imposed on corrupt actors and their enablers.

AI systems are also at risk of being exploited by corrupt actors to more closely mimic legitimate economic transactions in order to bypass detection by obliged entities or the authorities. As AI systems are increasingly adopted to support the detection of suspicious or anomalous transaction patterns at banks,[24] corrupt actors could use AI to analyse legitimate transaction patterns in order to generate a template for corrupt transactions, which is less likely to be identified by an AI monitoring system. These transaction patterns could be even more difficult to spot by obliged entities or the authorities than those conventionally used by corrupt actors.

USE OF AI TO MANIPULATE DISCOURSE DURING ELECTORAL CAMPAIGNS

People are increasingly linking AI to disinformation, with over 60 per cent of respondents from a global survey believing AI can create realistic fake news articles and images.[25] Some survey data conducted among EU states has shown that 83 per cent of respondents believe fake news poses a threat to democracy, especially concerning intentional disinformation targeting elections and immigration policies.[26] In 2024, the World Economic Forum ranked "AI-generated misinformation and disinformation" as the second most likely risk that could cause a "crisis on a global scale".[27]

AI tools can be used to customise disinformation campaigns targetting specific demographics, regions or individual voters, thereby enhancing their effectiveness, especially in electoral contexts.[28] As AI grows more powerful, its ability to influence voter decisions intensifies. For example, AI-generated deepfakes or manipulated media can make political candidates appear to say or do things they never did. [29,30] Fake news not only spreads incorrect information, but can also be deliberately weaponised with malicious intent – for example, to discredit electoral opponents through false corruption allegations, or to undermine journalists who accurately report cases of corruption. When such tools are deployed by state authorities, public officials or other individuals entrusted with power, this manipulation amounts to corruption, through use of AI.

USE OF AI TO WEAKEN SOCIAL ACCOUNTABILITY, OPPOSITION GROUPS AND DISSENT

Given the definition of corrupt use of AI as the intended abuse of AI systems for the private benefit of entrusted powerholders, any intentional use of AI systems to weaken social accountability, opposition groups or dissent also constitutes corrupt use of AI. Corruption thrives when it faces no opposition from those impacted by it. Integrity and accountability rely on the ability of the public, civil society, whistleblowers and the media to stand up for what is right, expose abuses of power, pursue remedies and trigger political change.[31]

Governments and other entities can deploy AI-driven surveillance to monitor and suppress dissenting voices, using facial recognition and data analysis to track activists and critics. AI algorithms can also be used to manipulate social media, spread disinformation, and create echo chambers to discredit opposition and sway public opinion.[32] At least 75 out of 176 countries globally are actively employing AI technologies for surveillance, frequently using them for smart policing and automated facial-recognition programmes.[33] Media reports have shown cases of large producers and exporters of AI-powered surveillance technology helping governments spy on political opponents.[34]

AI and data-driven mechanisms are widely used for controlling social media feeds, particularly for content moderation, sorting and generation. For example, the intentional use of AI to filter content may create content bubbles, specifically targeting users with an inherent political bias. This limits exposure to diverse viewpoints, and presents challenges to media pluralism.[35] Some states also deliberately employ AI to limit press freedom and deprioritise specific content, posing a significant risk to journalists' exposure online and distorting the freedom to seek, receive, publish and disseminate information concerning corruption.[36]

GOOD PRACTICE AND EXISTING STANDARDS ON AI INTEGRITY

In November 2021, UNESCO's 193 Member States adopted the first normative global instrument on the ethics of AI. The *UNESCO Recommendation on the Ethics of Artificial Intelligence* includes four overarching values, 10 principles and 11 policy areas with concrete proposals.[37] Under the Principle of Transparency and Explainability, the recommendation states that "greater transparency… allows for public scrutiny that can decrease corruption and discrimination and can also help detect and prevent negative impacts on human rights". It further notes that "explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should aim to be understandable and traceable".[38]

To date, the most comprehensive and legally binding international instrument regulating the use of AI by both public and private actors with a view to protecting rule of law, democracy and human rights is the *Council of Europe Framework Convention on AI*. The Convention acknowledges that there are serious risks and perils arising from certain activities within the AI lifecycle. These include discrimination in a variety of contexts, gender inequality, the undermining of democratic processes, the impairment of human dignity or individual autonomy, and the misuses of AI systems by some states for repressive purposes, in violation of international human rights law.[39]

The Organisation for Economic Cooperation and Development *(OECD) Recommendation on Artificial Intelligence* is an intergovernmental standard advocating for innovative and trustworthy AI that upholds human rights and democratic values. The recommendation also calls on AI actors to address risks arising from AI use outside the intended purpose, or from intentional misuse. This closely aligns with Transparency International's definition of corrupt AI use. The OECD Recommendation also considers AI use for disinformation and misinformation, and makes relevant recommendations to promote principles such as transparency, explainability and accountability.[40]

The UN Advisory Body report *Governing AI for Humanity* lays out five guiding principles for how AI should be governed. Guiding Principle 5 states that AI governance should be anchored in the UN Charter, international human rights law, and other agreed international commitments, such as the Sustainable Development Goals.[41] The principle states that the UN is positioned to consider AI's impact on a variety of global economic, social, health, security and cultural conditions, all grounded in the need to maintain universal respect for, and enforcement of, human rights and the rule of law.

The UN Global Digital Compact provides a framework and blueprint for the international governance of AI "for the benefit of humanity".[42] The compact's core ethical principles for AI emphasise transparency, accountability and fairness. It advocates respect for human rights, equitable distribution of AI benefits, and safeguards against harms. The principles stress combatting disinformation, ensuring authenticity of AI-generated content, and promoting open-source innovation to foster a safe, ethical digital environment.

At the European Union (EU) level, the AI Act is the most ambitious and comprehensive supranational normative framework regulating uses of AI. Article 1 notes the need to ensure a high level of protection of democracy and the rule of law against the harmful effects of AI systems.[43] While the Act covers intentional misuse of AI systems, public-sector misuse is only mentioned in relation to abuses by law enforcement or the use of social scoring systems.[44]

The World Health Organization has issued AI ethics and governance guidance for large multi-modal models, recommending that governments introduce mandatory post-release auditing and impact assessments – including data protection and human rights – by independent third parties when such a model is deployed on a large scale.[45]

At the national level, good examples include the creation of the Spanish Agency for the Supervision of Artificial Intelligence, the first AI regulatory body in the EU.[46] The UK's Data Ethics Framework guides ethical data use in the public sector, emphasising transparency and fairness, while Canada's Directive on Automated Decision-Making sets clear guidelines for AI use in government services, focusing on accountability and transparency.[47]

At the local level, nine European cities, collaborating through the Eurocities network, have developed an open source "data schema" for algorithm registers, to promote transparent and ethical use of AI in urban governance. This initiative, led by Barcelona, Bologna, Brussels Capital Region, Eindhoven, Mannheim, Rotterdam and Sofia, builds on the pioneering efforts of Amsterdam and Helsinki in creating AI registers. The data schema provides common guidelines for collecting information on algorithms used by cities, aiming to prevent data misuse and create an interoperable model that can be easily adopted by other municipalities, thus setting a standard for responsible AI use in European city administrations.[48]

RECOMMENDATIONS

As AI technology evolves, governments, industry and civil society should anticipate future challenges and regulations. AI systems blend technical infrastructure with the social contexts in which they are designed, developed and deployed. Accountability for these systems therefore requires transparency of their design, deployment and intended use. As AI continues to advance, more stringent and specific regulations are expected, particularly in critical areas such as health care and criminal justice, where transparency and explainability will be paramount.[49]

RECOMMENDATIONS TO GOVERNMENTS

1. Governments should show full commitment to addressing corrupt uses of AI by taking an inclusive and strategic approach to AI regulation. At a minimum, this should include the participatory drafting and adoption of a national AI strategy that includes specific objectives and activities targeted at preventing corrupt uses of AI and promoting AI integrity.
2. Governments should deliver on AI transparency and explainability by collecting and disclosing key information regarding AI systems under their management or oversight.
    i. At a minimum, governments should establish a public AI register containing information related to the use of AI across public-sector institutions.
    ii. Governments should also strive to establish AI transparency standards for these registers, which should be structured in an open data format and include details such as algorithm name, description, supplier information and channels for reporting discrepancies.
3. At a minimum, governments should ensure that when deploying AI systems for automated decision making, due consideration is given to balancing efficiency and explainability. This could mean that for certain policy areas where explainability of decisions is paramount – for example, in criminal justice or health care – the choice of models and systems should reflect such aims.
4. Governments should observe high standards of public procurement integrity and transparency when commissioning AI systems.
    i. At a minimum, contracting authorities should publish their plans for establishing or procuring AI systems well in advance, providing basic information on the intended purposes of these systems. This ensures transparency and allows for public oversight before implementation.
    ii. Contracting authorities should also require potential suppliers to demonstrate transparency – for example, by making technical details, including source code, available to independent experts for periodic performance inspection. This enhances accountability and ensures the systems function as intended.
    iii. At a minimum, contracting authorities should also have in place impact assessments and audit systems which reduce and mitigate risks of the intentional misuse of such systems for private gain.
5. Governments and oversight bodies should update their election integrity frameworks and monitoring tools to minimise the abuse and impact of AI tools and systems by state and political actors to spread disinformation, manipulate public opinion, and undermine the integrity of elections. At a minimum, the legal framework must define, restrict and provide oversight mechanisms against manipulative behaviours associated with AI-generated political content.
    i. Political parties, candidates, and third parties participating in campaigns should be provided with guidelines on the appropriate use of AI tools and methods as well as clear redlines on misuse or

manipulative behaviour. Social media platforms should also be provided with these redlines for misuse as guidance for their oversight and content moderation. A standard list of such redlines for misuse and manipulative behaviour can be found in the EU Code of Conduct on Dinsinformation.[50]

ii. All political advertising generated or disseminated through AI tools should contain a disclaimer disclosing what technology was used to generate or disseminate the material, to enable voters to make informed decisions. Watermarking and content provenance protocols such as C2PA should also be encouraged to ensure content traceability.[51]

iii. Introduce standardised benchmarks and guidance for oversight bodies and social media platforms to strengthen their monitoring, detection and enforcement capacities.

6. The legal framework should provide for an empowered agency to supervise and regulate the use of AI for the common good.

i. At a minimum, the legal framework should provide for the establishment of an independent agency, or a division within an independent agency, mandated to promote responsible AI, including the prevention of corrupt AI use. This should include AI literacy programmes to improve understanding among public officials, regulators and the general public, with particular emphasis on electoral integrity and applications in electoral contexts.

ii. Furthermore, the agency should make recommendations across public users of AI systems on data minimisation, necessity and proportionality.

iii. The agency or division should have the independence, budget, capacity and resources to fulfil its mandate, and should establish cooperation and collaboration modalities with other relevant institutions, domestically and internationally.

7. The legal framework should introduce measures for redress against corrupt use of AI. At a minimum, such measures should include provision for individuals to appeal a decision made by an AI system; mechanisms to monitor and override decisions made by an AI system; safe channels for reporting suspected misuse of AI systems; a procedure to investigate reports, and protection for whistleblowers.

RECOMMENDATIONS TO AI PROVIDERS AND COMPANIES

1. AI system providers should be legally required to conduct customer due diligence when selling to public-sector clients, particularly in contexts with weak democratic checks and balances.

2. At a minimum, AI providers should conduct know-your-customer checks on their clients, and conduct a risk assessment on the potential for corrupt misuse of the AI system. The due diligence should include creating case profiles for intended and risk-appropriate uses of the AI system.

3. At a minimum, AI providers should implement internal systems to monitor their clients' use of AI and detect potential misuse. Any suspicion of misuse should be reported to the appropriate government authorities, such as the anti-corruption or AI supervision authority. In case of non-response or corrupt involvement of government authorities, the AI provider should withdraw the licence from the client.

4. AI providers should be required to communicate attempts of misuse of their AI systems for corrupt purposes with the authorities and, where possible, with the public, providing detailed information on the type of misuse and remedial actions taken. Examples of such misuse may include significant deployment of their systems to generate or disseminate disinformation, manipulate public opinion, curtail civil and political rights, automate discriminatory decisions, or otherwise undermine democratic processes or public integrity. In the case of abuses of AI systems on online platforms, they should develop measures to warn users when such practices are detected.

5. AI providers should take a participatory, transparent and collaborative approach to data sharing with the authorities, civil society, academia and the wider public, to promote accountability and generate trust.

i. At a minimum, AI providers should have in place mechanisms to share with the authorities any information relating to the misuse of their systems for corrupt purposes, including streamlined processes to respond to requests for data from the authorities.

ii. At a minimum, AI providers should share data and collaborate with civil society to conduct algorithm audits and provide participatory oversight of the deployment of AI systems that are exposed to corruption risks.

6. Obliged entities that detect the misuse of their AI systems for laundering the proceeds of corruption should go beyond their legal obligations to report suspicious transactions and follow best practice by reporting misuses of their systems to their respective AI providers, to prevent future misuse.
7. To meet future challenges, AI providers should share data on patterns of systems use and misuse; data and algorithm audit reports, and other relevant data with appropriate government authorities, civil society and academia, to test for potential misuse of their systems and conduct horizon scanning for likely future developments.
8. Businesses which perform a public function, or provide a public service as defined by law, should follow all recommendations and requirements placed on public-sector entities or public enterprises in relation to preventing corrupt use of AI.

RECOMMENDATIONS TO CIVIL SOCIETY ORGANISATIONS

1. To effectively address challenges presented by the corrupt use of AI, civil society organisations (CSOs) working on issues of governance, integrity, technology and their intersection should, at a minimum:
   i. Build internal capacities to be able to detect and counter corruption and integrity risks as they relate to the use of AI systems in the public sector.
   ii. Partner with experts, data scientists and practitioners to ensure that CSO frameworks remain up to date in the face of rapid technological shifts.
2. CSOs should seek to actively participate in policy discussions and consultations related to AI governance, including by providing input on draft regulations, and advocating for robust safeguards against the corrupt use of AI.
   i. Through awareness-raising and capacity-building initiatives, CSOs should promote knowledge among targeted stakeholders and the public about potential risks and mitigation strategies for the corrupt use of AI in governance.
   ii. Review existing or proposed legal frameworks regulating AI use, and assess whether they meet domestic, regional or international standards and commitments, such as the UN Global Digital Compact and the EU AI Act.

ENDNOTES

[1] Competition Policy International, "New Report Says AI Regulations Lag Behind Industry Advances", accessed 13 May 2025, https://www.competitionpolicyinternational.com/new-report-says-ai-regulations-lag-behind-industry-advances/.
[2] Brookings Institution, "The Three Challenges of AI Regulation", accessed 13 May 2025, https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/.
[3] Transparency International, *Holding Power to Account: A Global Strategy 2021-2030*, 2021, Strategy2030_Brochure-final_15022021.pdf.
[4] Transparency International, *Holding Power to Account: A Global Strategy 2021-2030*.
[5] Freedom House, "Repressive Power of Artificial Intelligence", 2023, https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence.
[6] Transparency International, "What is corruption?" Accessed 13 May 2025, https://www.transparency.org/en/what-is-corruption.
[7] Transparency International, *Holding Power to Account: A Global Strategy Against Corruption 2021-2030*.
[8] SpringerLink, "Artificial Intelligence for Decision Making in the Era of Big Data", Accessed 13 May 2025, https://link.springer.com/article/10.1007/s13347-022-00577-5.
[9] CORE, "Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda", 2019, Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda - ScienceDirect.
[10] For example, in England and Wales. ResearchGate, "From AI to Admissible Implications of Computer-Generated Evidence in Commonwealth Courts", 2024, https://www.researchgate.net/publication/384323910_From_AI_to_Admissible_Implications_of_Computer-Generated_Evidence_in_Commonwealth_Courts; *The Guardian*, "Update Law on Computer Evidence to Avoid Horizon Repeat, Ministers Urged", 12 January 2024, https://www.theguardian.com/uk-news/2024/jan/12/update-law-on-computer-evidence-to-avoid-horizon-repeat-ministers-urged.
[11] Partnership on AI, "Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System", 2021, https://partnershiponai.org/wp-content/uploads/2021/08/Report-on-Algorithmic-Risk-Assessment-Tools.pdf.
[12] At least 75 out of 176 countries globally are actively using AI technologies for surveillance purposes, often for smart policing and automated facial-recognition programmes. Carnegie Endowment for International Peace, "The Global Expansion of AI Surveillance", 2019, https://carnegieendowment.org/research/2019/09/the-global-expansion-of-ai-surveillance?lang=en.
[13] National Endowment for Democracy, "The Digital Battlefield for Democratic Principles", 2023, https://www.ned.org/wp-content/uploads/2023/03/NED_Forum-The-Digital-Battlefield-for-Democratic-Principles.pdf.
[14] World Economic Forum, "AI Procurement in a Box: AI Government Procurement Guidelines", 2020, https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf.
[15] Oxford Pathways Commission, "Are Emerging Technologies Helping Win the Fight Against Corruption in Developing Countries?", 2019, https://pathwayscommission.bsg.ox.ac.uk/sites/default/files/2019-09/are_emerging_technologies_helping_win_the_fight_against_corruption_in_developing_countries.pdf.
[16] OECD, *Anticorruption and Integrity Outlook 2024*, 2024, Artificial intelligence | OECD.
[17] UK Government, "Review into Bias in Algorithmic Decision-Making", 2021, https://assets.publishing.service.gov.uk/media/60142096d3bf7f70ba377b20/Review_into_bias_in_algorithmic_decision-making.pdf.
[18] The Fast Mode, "Using Bias Intentionally in Artificial Intelligence", 2024, https://www.thefastmode.com/expert-opinion/33425-using-bias-intentionally-in-artificial-intelligence.
[19] Transparency International, "The Corruption Risks of Artificial Intelligence", 2022, https://knowledgehub.transparency.org/assets/uploads/kproducts/The-Corruption-Risks-of-Artificial-Intelligence.pdf.
[20] Transparency International, "The Corruption Risks of Artificial Intelligence".
[21] Economist Impact, "AI in Financial Services", 2024, https://impact.economist.com/perspectives/sites/default/files/aiinfinancialservices.pdf; https://www.bis.org/publ/arpdf/ar2024e3.htm.
[22] OECD, Anti-Corruption and Integrity Outlook 2024, 2024, https://www.oecd-ilibrary.org/governance/anti-corruption-and-integrity-outlook-2024_968587cd-en.
[23] For example, the extensive use of fraudulent documents in the South African state capture case. Amabhungane Centre for Investigative Journalism, "GuptaLeaks: Meet the Money Launderers", 2017, https://amabhungane.org/guptaleaks-meet-the-money-launderers/.
[24] O. Berkan, D. Cetinkaya, F. Adedoyin, M. Budka, G. Aksu and H. Dogan, "Transaction Monitoring in Anti-Money Laundering: A Qualitative Analysis and Points of View from Industry", 2024, https://www.sciencedirect.com/science/article/pii/S0167739X24002607.

[25] Politico, "How People View AI and Disinformation: Perceptions Around Elections", 2024, https://www.politico.eu/article/people-view-ai-disinformation-perception-elections-charts-openai-chatgpt/.
[26] European Commission, *Eurobarometer Survey on Trust in Institutions*, 2024, https://europa.eu/eurobarometer/surveys/detail/2183.
[27] World Economic Forum, *Global Risks Report 2024,* https://www.weforum.org/publications/global-risks-report-2024/.
[28] C. Yu, "How Will AI Steal Our Elections?", Center for Open Science, 2024, https://ideas.repec.org/p/osf/osfxxx/un7ev.html.
[29] European Commission, *Annex I: Literature Review on Disinformation and AI,* 2022, https://commission.europa.eu/system/files/2022-12/Annex%20I_LiteratureReview_20210319_clean_dsj_v3.0_a.pdf.
[30] Brennan Center for Justice, "How AI Puts Elections at Risk—And the Needed Safeguards", 2023, https://www.brennancenter.org/our-work/analysis-opinion/how-ai-puts-elections-risk-and-needed-safeguards.
[31] Transparency International, *Holding Power to Account: A Global Strategy Against Corruption 2021-2030*.
[32] An active and prolific, but ultimately low-impact, cross-platform political spam network in China boosted attacks on the Hong Kong protesters by using hijacked or fake accounts on YouTube, Twitter and Facebook. Graphika, "Spamouflage: Prolific but Low-Impact Cross-Platform Political Spam Network", 2021, https://graphika.com/reports/spamouflage.
[33] S. Feldstein, "The Global Expansion of AI Surveillance*",* Carnegie Endowment for International Peace, 2019, https://carnegie-production-assets.s3.amazonaws.com/static/files/WP-Feldstein-AISurveillance_final1.pdf.
[34] J. Parkinson and D. Chin, "Huawei Technicians Helped African Governments Spy on Political Opponents", *The Wall Street Journal*, 15 August 2019, https://www.wsj.com/articles/huawei-technicians-helped-african-governments-spy-on-political-opponents-11565793017.
[35] Organization for Security and Co-operation in Europe (OSCE), Comprehensive Review of AI Use in Electoral Context*,* 2021, https://www.osce.org/files/f/documents/4/5/472488.pdf.
[36] OSCE, Comprehensive Review of AI Use in Electoral Context.
[37] UNESCO, "*Recommendation on the Ethics of Artificial Intelligence"*, 2021, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.
[38] UNESCO, "*Recommendation on the Ethics of Artificial Intelligence"*, 2022, Sections 39 and 40, p. 22, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.
[39] Council of Europe, "Explanatory Report to the CETS No.225", 2024, https://rm.coe.int/1680afae67.
[40] OECD, "Legal Instruments, *Recommendation on Public Integrity",* 2019, https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf.
[41] UN Advisory Body on Artifical Intelligence, "*Governing AI for humanity : final report,"* September 2024, *https://digitallibrary.un.org/record/4062495?v=pdf*.
[42] United Nations, "*Global Digital Compact",* September 2024, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.
[43] European Parliament, "Corrigendum to the European Parliament Resolution on Artificial Intelligence in the Digital Age", April 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf.
[44] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf European Parliament, "Corrigendum to the European Parliament Resolution on Artificial Intelligence in the Digital Age".
[45] World Health Organization, "Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models," Geneva: WHO, 2024, https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models.
[46] Spanish Agency for the Supervision of Artificial Intelligence, "Spanish Agency for the Supervision of Artificial Intelligence (AESIA)," Government of Spain. Accessed June 5, 2025. https://aesia.digital.gob.es/en/es.
[47] For the UK, UK Government, "Data Ethics Framework," 2020, https://www.gov.uk/government/publications/data-ethics-framework; for Canada Government of Canada, "*Digital Government Strategy*," 2020, https://www.canada.ca/en/government/system/digital-government/digital-government-strategy.html.
[48] Eurocities, "Nine Cities Set Standards for the Transparent Use of Artificial Intelligence", 19 January 2023, https://eurocities.eu/latest/nine-cities-set-standards-for-the-transparent-use-of-artificial-intelligence/.
[49] According to the European Data Protection Supervisor, "Explainable Artificial Intelligence (XAI) is the ability of AI systems to provide clear and understandable explanations for their actions and decisions. Its central goal is to make the behaviour of these systems understandable to humans by elucidating the underlying mechanisms of their decision-making processes." European Data Protection Supervisor, "TechDispatch: Explainable Artificial Intelligence (XAI)", 16 November 2023, https://www.edps.europa.eu/system/files/2023-11/23-11-16_techdispatch_xai_en.pdf.
[50] A standard list of such redlines for misuse can be found in Commitment 14 in the EU Code of Conduct on Dinsinformation, European Commission, "EU Code of Conduct on Dinsinformation," 2025, https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation.
[51] Coalition for Content Provenance and Authenticity (C2PA), "C2PA: Coalition for Content Provenance and Authenticity," 2024, https://c2pa.org.

ENGAGE

Follow us, share your views and discuss corruption with people from around the world on social media.

Facebook: /transparencyinternational
Twitter/X: @anticorruption
LinkedIn: @transparency-international
Instagram: @Transparency_International
YouTube: @TransparencyIntl

LEARN

Visit our website to learn more about our work in more than 100 countries and sign up for the latest news in the fight against corruption.

transparency.org

DONATE

Your donation will help us provide support to thousands of victims of corruption, develop new tools and research, and hold governments and businesses to their promises. We want to build a fairer, more just world. With your help, we can.

transparency.org/donate

Transparency International
International Secretariat
Alt-Moabit 96, 10559 Berlin, Germany

Phone: +49 30 34 38 200

ti@transparency.org
www.transparency.org

Blog: transparency.org/blog
Facebook: /transparencyinternational
Twitter/X: @anticorruption
LinkedIn: @transparency-international
Instagram: @Transparency_International
YouTube: @TransparencyIntl