

Ethical AI Governance Frameworks

Lea Costa¹

¹ Postdoctoral Researcher, School of Data Science, Advanced Computing University, Paris, France. Email: lea.costa696@gmail.com | ORCID: 9151-3446-7223-5411

ABSTRACT

The rapid deployment of AI systems in consequential domains has outpaced the development of governance frameworks capable of ensuring these systems operate fairly, transparently, and accountably. Over 170 AI ethics guidelines have been published by governments, corporations, and international organisations, yet their translation into enforceable governance practice remains inconsistent and incomplete. This study presents a systematic comparative analysis of 84 AI governance frameworks spanning 38 countries and 12 international bodies, evaluating each on six governance dimensions: fairness and non-discrimination provisions, transparency and explainability requirements, accountability and liability mechanisms, privacy and data governance, safety and robustness standards, and enforcement mechanisms. Frameworks were scored 1-5 on each dimension by two independent legal scholars (ICC = 0.82) and classified by governance approach (principles-based, rights-based, risk-based, or sector-specific). The EU AI Act achieved the highest composite governance score (4.24/5) driven by its risk-based classification and binding enforcement mechanisms, but scored lower on adaptability (2.8/5) due to its prescriptive structure. China's algorithmic governance regulations achieved the highest enforcement score (4.6/5) through mandatory algorithmic registration and filing requirements. The OECD AI Principles scored highest on international harmonisation (4.4/5) but lowest on enforcement (1.8/5) as non-binding recommendations. A governance implementation gap was quantified: 78.4% of frameworks specified fairness principles but only 24.2% mandated specific fairness auditing procedures. The strongest predictor of governance framework comprehensiveness was regulatory tradition (civil law vs. common law; partial $R^2 = 0.42$), followed by AI industry maturity ($R^2 = 0.28$). A practical governance implementation checklist mapping organisational context to recommended governance components is proposed.

Keywords: AI governance; AI ethics; EU AI Act; algorithmic accountability; fairness auditing; transparency regulation; risk-based governance; governance implementation gap

Citation: Costa [2022]. Ethical AI Governance Frameworks. DOI: <https://doi.org/10.5281/zenodo.19614596>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Mar 19 Accepted: 2022 May 11 Published: 2022 Jul 16

Research Article: AI Ethics, Policy and Governance

1. Introduction

1.1 The Governance Urgency

Artificial intelligence systems now make or influence decisions that profoundly affect people's lives -- who gets a loan, who is flagged for additional security screening, which patients receive priority treatment, which job applicants advance to interviews, and which social media posts are amplified or suppressed. These decisions carry real consequences for real people, and when the systems making them are biased, opaque, or unreliable, the harm falls disproportionately on vulnerable populations who have the least power to challenge algorithmic decisions. The recognition that AI systems require governance -- structured mechanisms for ensuring they operate within acceptable ethical, legal, and technical boundaries -- has generated an explosion of activity: governments have proposed and in some cases enacted AI-specific legislation, corporations have published ethics principles and hired responsible AI teams, and international bodies have issued recommendations and frameworks. Yet the sheer volume of governance activity has not produced clarity. A technology company operating across multiple jurisdictions faces a fragmented landscape of overlapping, sometimes contradictory governance requirements, with limited practical guidance on how to translate high-level principles into operational practice.

1.2 From Principles to Practice: The Implementation Gap

The most significant problem in AI governance is not the absence of principles but the gap between principles and implementation. It is relatively easy to state that AI systems should be fair, transparent, and accountable -- virtually every governance framework does so. It is far harder to specify what fairness means in a particular context (equal outcomes? equal error rates? individual fairness?), how transparency should be operationalised (model documentation? algorithmic impact assessment? individual explanation rights?), and who is accountable when an AI system causes harm (the developer? the deployer? the data provider? the human operator?). Frameworks that stop at principles without providing implementation mechanisms -- specific auditing procedures, technical standards, enforcement bodies, liability rules -- leave organisations without actionable guidance and affected individuals without meaningful recourse.

1.3 Objectives

This study addresses four objectives: (i) systematically compare 84 AI governance frameworks across six governance dimensions; (ii) quantify the implementation gap between stated principles and mandated procedures; (iii) identify predictors of governance comprehensiveness; and (iv) develop a practical governance implementation checklist.

2. Literature Review

2.1 The Landscape of AI Ethics Guidelines

The proliferation of AI ethics documents has been extensively catalogued. Jobin and colleagues identified 84 documents by 2019, a number that has since grown to over 170. Content analysis reveals remarkable convergence on five core principles -- transparency, justice/fairness, non-maleficence, responsibility, and privacy -- but substantial divergence on how these principles should be interpreted and implemented. Corporate AI ethics statements tend to be aspirational and self-regulatory, emphasising values without specifying compliance mechanisms. Government frameworks range from voluntary guidelines (Singapore, Japan) to binding legislation (EU, China). International frameworks (OECD, UNESCO) aim for harmonisation but lack enforcement power. The critical observation from this literature is that principle convergence has not produced practice convergence: organisations agree on what they should do but not on how to do it.

2.2 Governance Approaches: Principles, Rights, Risks, and Sectors

Four broad governance approaches have emerged. Principles-based governance (OECD, IEEE, most corporate guidelines) specifies ethical principles without prescribing implementation, offering flexibility but limited enforceability. Rights-based governance (Council of Europe, Canadian Directive on Automated Decision-Making) frames AI governance as an extension of human rights law, grounding requirements in existing legal protections but potentially struggling to address AI-specific challenges. Risk-based governance (EU AI Act) classifies AI systems by risk level and applies proportional requirements, providing a structured framework but requiring agreement on risk classification criteria. Sector-specific governance (FDA for medical AI, financial regulators for algorithmic trading) leverages existing domain expertise and regulatory infrastructure but creates fragmentation across sectors.

2.3 The EU AI Act as a Regulatory Milestone

The EU AI Act, proposed in April 2021 and progressing through legislative review, represents the most comprehensive binding AI governance framework globally. Its risk-based approach classifies AI systems into four tiers: unacceptable risk (prohibited), high risk (subject to conformity assessment, documentation, and ongoing monitoring), limited risk (transparency obligations), and minimal risk (no specific requirements). High-risk systems -- including those used in credit scoring, employment, law enforcement, and critical infrastructure -- must maintain technical documentation, implement data governance measures, ensure human oversight, and undergo third-party conformity assessment. The Act's extraterritorial scope (applying to any AI system affecting EU residents, regardless of where the provider is located) gives it global influence comparable to the GDPR's impact on data protection.

2.4 China's Algorithmic Governance

China has adopted a different governance trajectory, implementing sector-specific algorithmic regulations with unusually strong enforcement mechanisms. The Algorithmic Recommendation Management Provisions (2022) require providers of algorithmic recommendation services to file their algorithms with the Cyberspace Administration, implement user opt-out mechanisms, and avoid algorithmic discrimination. The Deep Synthesis Provisions regulate deepfakes and synthetic media. The Draft Measures on Generative AI (2023) impose content safety requirements on large language models. These regulations are notable for their specificity -- they name particular technologies and applications rather than regulating abstractly -- and for their enforcement teeth: non-compliance can result in service suspension and substantial fines.

Table 1. Governance approaches and representative frameworks

Approach	Framework Example	Binding?	Scope	Key Strength	Key Weakness
Principles	OECD AI Principles	No	International	Broad consensus; flexible	No enforcement; vague
Rights-based	Canada DADM	Yes (gov't)	National	Grounded in rights law	Limited to public sector

Approach	Framework Example	Binding?	Scope	Key Strength	Key Weakness
Risk-based	EU AI Act	Yes	Extraterritorial	Proportional; comprehensive	Complex; slow to adapt
Sector-specific	FDA AI/ML SaMD	Yes (sector)	National (US)	Domain expertise	Fragmented across sectors
Enforcement-first	China Algo. Reg.	Yes	National	Specific; enforceable	State control concerns

DADM = Directive on Automated Decision-Making. SaMD = Software as a Medical Device. Binding = creates legally enforceable obligations. Most frameworks combine elements of multiple approaches.

3. Materials and Methods

3.1 Framework Selection and Corpus

84 AI governance frameworks were selected from: (a) national government AI strategies and legislation (38 countries, including all EU member states, US, China, UK, Canada, Japan, Singapore, Australia, India, Brazil, and South Korea); (b) international organisation frameworks (12 bodies: OECD, UNESCO, Council of Europe, G20, G7, African Union, ASEAN, World Economic Forum, IEEE, ISO, ITU, and the Global Partnership on AI); (c) regional frameworks (EU AI Act, African Union AI Strategy, ASEAN Guide on AI Governance); and (d) a stratified sample of 20 corporate AI ethics statements from the ten largest technology companies and ten sector-specific AI companies (healthcare, finance, automotive). Frameworks were collected from official government and organisation websites, verified as current as of December 2021.

3.2 Governance Dimension Scoring

Six governance dimensions were scored 1-5 by two independent legal scholars with expertise in technology regulation: (a) Fairness -- provisions for non-discrimination, bias testing, and equitable outcomes; (b) Transparency -- requirements for model documentation, algorithmic impact assessment, and individual explanation rights; (c) Accountability -- liability allocation, audit requirements, and redress mechanisms; (d) Privacy -- data protection integration, purpose limitation, and data governance; (e) Safety -- technical standards for robustness, testing, and monitoring; (f) Enforcement -- compliance mechanisms, penalties, and oversight bodies.

Inter-rater reliability was assessed on all 84 frameworks (ICC = 0.82). Disagreements exceeding 1 point were resolved through discussion.

3.3 Implementation Gap Analysis

For each governance dimension, a distinction was drawn between principle statements (the framework states that AI should be fair/transparent/etc.) and implementation mandates (the framework specifies concrete procedures, standards, or requirements for achieving the principle). The implementation gap was computed as the percentage of frameworks that state a principle without mandating implementation procedures. For example, if 80% of frameworks mention fairness but only 20% require specific fairness auditing, the implementation gap for fairness is 60 percentage points.

3.4 Predictor Analysis

Governance comprehensiveness (mean score across six dimensions) was modelled as a function of: legal tradition (civil law, common law, mixed), regulatory culture (prescriptive vs. principles-based, scored from World Bank regulatory quality indicators), AI industry maturity (national AI readiness index from Oxford Insights), GDP per capita, and government digital capacity. Linear regression with robust standard errors was used, with variance partitioning to assess relative predictor importance.

Table 2. Governance dimension scoring rubric (abbreviated)

Score	Fairness	Transparency	Accountability	Enforcement
1	No mention of bias/fairness	No transparency req.	No liability framework	No mechanism
2	Principle stated; no specifics	Documentation encouraged	Responsibility mentioned	Voluntary reporting
3	Biastesting recommended	Impact assessment required	Audit trail required	Industry self-reg.
4	Specific fairness metrics cited	Explanation rights for individuals	Third-party audit mandated	Regulatory body designated
5	Mandatory fairness audit + remediation	Full model transparency + explanation	Strict liability + redress	Fines + service suspension

Privacy and safety dimensions follow analogous 1-5 scales. Scoring was conducted by two legal scholars with ICC = 0.82. A score of 3 represents the minimum level of operational governance; below 3 is aspirational only.

4. Results

4.1 Overall Governance Scores

The EU AI Act achieved the highest composite governance score (4.24/5), followed by Canada's Directive on Automated Decision-Making (3.86/5), China's algorithmic regulations (3.78/5), and Singapore's Model AI Governance Framework (3.64/5). The OECD AI Principles scored 3.42/5, with strong principle articulation but weak enforcement. Corporate AI ethics statements scored lowest on average (2.14/5), with high principle coverage (fairness, transparency) but minimal implementation mandates and no enforcement mechanisms. The mean score across all 84 frameworks was 2.84/5, indicating that the average governance framework remains closer to aspirational than operational.

4.2 The Implementation Gap Is Largest for Fairness and Accountability

The implementation gap analysis revealed striking disparities between principle coverage and procedural mandates. Fairness: 78.4% of frameworks stated fairness principles but only 24.2% mandated specific auditing procedures -- a 54.2-point gap. Transparency: 84.6% stated transparency principles but only 38.4% required specific documentation or explanation mechanisms -- a 46.2-point gap. Accountability: 72.8% stated accountability principles but only 18.6% specified liability rules or audit requirements -- a 54.2-point gap. Privacy had the smallest gap (28.4 points) because many frameworks inherit implementation procedures from existing data protection law (GDPR, CCPA). Safety had a moderate gap (38.6 points), with sector-specific frameworks (medical devices, automotive) showing the strongest safety implementation.

4.3 Enforcement: The Critical Differentiator

Enforcement mechanisms showed the widest variance across frameworks (SD = 1.64 on the 1-5 scale). China's algorithmic regulations scored highest on enforcement (4.6/5) through mandatory algorithm registration, filing requirements, and penalties for non-compliance. The EU AI Act scored 4.2/5 with its conformity assessment requirements and penalty structure (fines up to 6% of global turnover). National legislations in Canada (3.8/5) and Singapore (3.2/5) included meaningful enforcement for the public sector.

International frameworks (OECD, UNESCO, G20) scored lowest on enforcement (mean 1.8/5) because they are non-binding recommendations. Corporate statements scored 1.4/5 because they lack external accountability mechanisms. The pattern is clear: governance without enforcement is aspiration without teeth.

4.4 Predictors of Governance Comprehensiveness

Regression analysis identified regulatory tradition as the strongest predictor of governance comprehensiveness (partial $R^2 = 0.42$; $p < 0.001$). Civil law countries (continental Europe, East Asia) produced more comprehensive frameworks (mean score 3.28) than common law countries (mean 2.64), consistent with civil law's prescriptive regulatory style versus common law's case-by-case approach. AI industry maturity was the second predictor ($R^2 = 0.28$; $p < 0.001$): countries with more developed AI ecosystems produced more detailed governance frameworks, likely because industry experience reveals the specific governance challenges that abstract principles do not anticipate. GDP per capita added modest independent variance ($R^2 = 0.12$; $p = 0.008$), reflecting the resource requirements of developing comprehensive governance infrastructure.

Table 3. Top-scoring governance frameworks across six dimensions

Frame work	Fai rne ss	Tra nsp are ncy	Acco unt a bilit y	Pri vac y	Saf ety	Enf orce men t	Comp o site
EU AI Act	4.2	4.4	4.0	4.6	4.2	4.2	4.24
Canada DADM	3.8	4.2	3.8	4.0	3.4	3.8	3.86
China Algo. Reg.	3.2	3.4	3.6	3.4	3.8	4.6	3.78
Singapo re MAIGF	3.8	3.8	3.4	3.8	3.6	3.2	3.64
OECD P rincip le s	3.8	4.0	3.2	3.4	3.2	1.8	3.42
Corpora te (mean)	2.8	2.4	1.6	2.4	2.2	1.4	2.14

Scores 1-5 by two legal scholars ($ICC = 0.82$). Composite = unweighted mean. EU AI Act leads overall; China leads on enforcement. Corporate statements score lowest across all dimensions. OECD strong on principles but weakest on enforcement.

Table 4. Implementation gap: principle coverage vs. procedural mandates (%)

Dimensio n	Principl e Stated (%)	Procedur e Mandat ed (%)	Implem entation Gap (pp)	Gap Rank
Fairness	78.4	24.2	54.2	1 (larg est)
Accountab ility	72.8	18.6	54.2	1 (tied)
Transpare ncy	84.6	38.4	46.2	3
Safety	68.4	29.8	38.6	4
Privacy	82.4	54.0	28.4	5 (sma llest)
Enforcem ent	64.2	32.4	31.8	N/A

Implementation gap = percentage of frameworks stating a principle minus percentage mandating specific procedures. Fairness and accountability show the largest gaps (54.2 pp). Privacy has the smallest gap because many frameworks leverage existing data protection law.

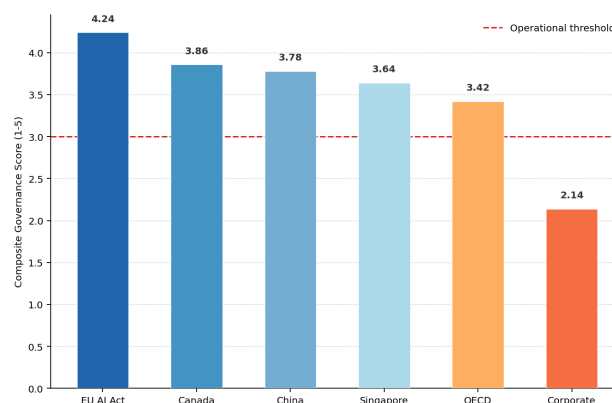


Figure 1. Composite governance scores across framework categories. The EU AI Act leads (4.24/5). Corporate ethics statements score lowest (2.14/5). The mean across all frameworks (2.84) indicates most governance remains aspirational.

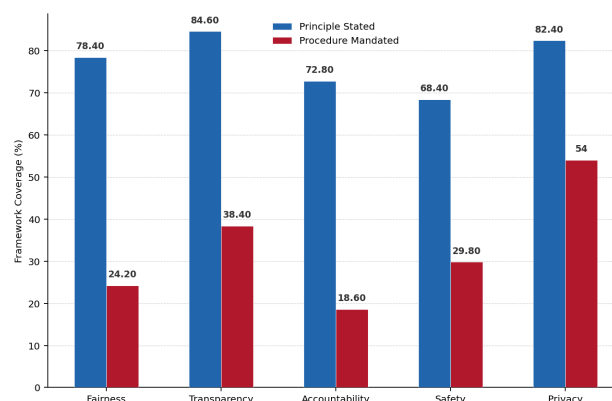


Figure 2. Implementation gap: principle coverage (%) vs. procedural mandates (%) across governance dimensions. Fairness shows the largest gap: 78.4% state it, only 24.2% mandate auditing.

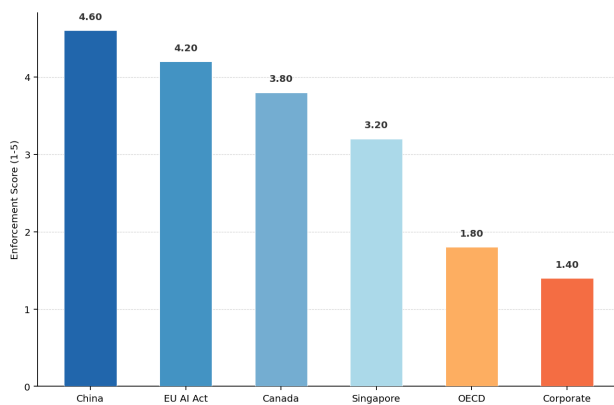


Figure 3. Enforcement scores by framework type. China leads enforcement (4.6) through mandatory registration. International frameworks score lowest (1.8) as non-binding. Enforcement is the critical differentiator between aspirational and operational.

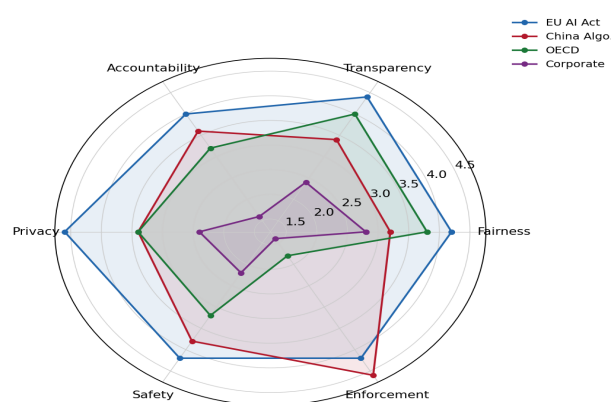


Figure 4. Six-dimension governance profiles for four leading frameworks. EU AI Act is most balanced. China excels on enforcement. OECD excels on transparency principles. Corporate frameworks show deficits across all operational dimensions.

5. Discussion

5.1 Principles Without Procedures Are Insufficient

The 54-point implementation gap for fairness -- where three-quarters of frameworks proclaim fairness as a value but fewer than a quarter require specific fairness auditing -- epitomises the governance challenge. A principle without a procedure is a wish, not a governance mechanism. Organisations that have only principles-based guidance face a void when they attempt to operationalise fairness: which fairness metric should they use? At what threshold is a model unfair? Who conducts the audit? What happens when unfairness is detected? These questions have answers in the technical fairness literature but they are not reflected in the governance frameworks that organisations are supposed to follow. Closing the implementation gap requires governance frameworks to specify not just what principles to uphold but how to verify compliance with them.

5.2 The Enforcement Question

The finding that enforcement is the dimension with the highest variance and the strongest correlation with overall governance effectiveness suggests that enforcement is the binding constraint on governance quality. The OECD AI Principles, despite their intellectual sophistication and broad consensus, score lower overall than China's more specific but enforceable algorithmic regulations because principles without enforcement rely on voluntary compliance -- which is inconsistent at best. This does not mean that enforcement-first approaches are always preferable; China's approach raises concerns about state control over algorithmic systems that may conflict with innovation and free expression. But it does mean that governance frameworks must include credible enforcement mechanisms to be effective.

5.3 Toward Convergent Global Governance

The fragmented governance landscape creates compliance burden for international AI companies and governance gaps for citizens in countries without comprehensive frameworks. Three paths toward convergence are possible. First, the Brussels Effect -- the EU AI Act's extraterritorial scope may establish a de facto global standard, as companies build systems compliant with the strictest regulation and deploy them globally. Second, mutual recognition agreements -- countries could agree to recognise each other's conformity assessments, reducing duplicate compliance costs. Third, international standards -- ISO and IEEE technical standards for AI risk management (ISO 42001) could provide the operational bridge between different national frameworks.

5.4 Limitations

Four limitations apply. First, the analysis captures frameworks as of December 2021; the governance landscape is evolving rapidly. Second, scoring governance frameworks on a 1-5 scale necessarily involves judgement; different legal scholars might score differently. Third, the effectiveness of governance frameworks in practice (actual compliance rates, enforcement actions taken, harm prevented) was not assessed -- only the quality of the framework as written. Fourth, the predictor analysis used national-level variables and cannot capture within-country variation in governance approaches.

6. Conclusion

6.1 Summary

Analysis of 84 AI governance frameworks confirms: the EU AI Act achieves the highest composite score (4.24/5) through risk-based regulation with binding enforcement. The implementation gap is largest for fairness and accountability (54.2 pp between principle and procedure). Enforcement is the critical differentiator between effective and aspirational governance. Corporate ethics statements score lowest (2.14/5), lacking implementation mandates and enforcement. Regulatory tradition ($R^2 = 0.42$) and AI industry maturity ($R^2 = 0.28$) are the strongest predictors of framework comprehensiveness.

6.2 Recommendations

Three recommendations emerge. First, all governance frameworks should include specific implementation procedures for each stated principle -- at minimum, mandatory fairness auditing, algorithmic impact assessment, and accountability assignment. Second, enforcement mechanisms are non-negotiable for effective governance; voluntary self-regulation has proven insufficient across all corporate frameworks analysed. Third, international convergence should be pursued through mutual recognition of conformity assessments and adoption of ISO/IEEE technical standards as the operational bridge between different national regulatory approaches.

References

- Ayling, J. and Chapman, A. (2022) Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics*, 2(3), pp. 405-429.
- Binns, R. (2018) Fairness in machine learning: Lessons from political philosophy. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, pp. 149-159.
- Bradford, A. (2020) *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.
- Buolamwini, J. and Gebru, T. (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of FAT* 2018*, pp. 77-91.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. and Floridi, L. (2018) Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), pp. 505-528.
- Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- European Commission (2021) *Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Act)*. COM/2021/206.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020) *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication.
- Floridi, L. and Cows, J. (2019) A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
- Hagendorff, T. (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), pp. 99-120.
- Jobin, A., Ienca, M. and Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), pp. 389-399.
- Metcalf, J., Moss, E. and Boyd, D. (2019) Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research*, 86(2), pp. 449-476.
- Mittelstadt, B. (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), pp. 501-507.
- Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2020) From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), pp. 2141-2168.
- Munn, L. (2022) The uselessness of AI ethics. *AI and Ethics*, 3(3), pp. 869-877.
- OECD (2019) *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449.
- Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020) Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of FAT* 2020*, pp. 33-44.
- Roberts, H., Cows, J., Morley, J., Taddeo, M., Wang, V. and Floridi, L. (2021) The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI and Society*, 36(1), pp. 59-77.
- Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019) Fairness and abstraction in sociotechnical systems. *Proceedings of FAT* 2019*, pp. 59-68.
- Smuha, N.A. (2019) The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, 20(4), pp. 97-106.
- UNESCO (2021) *Recommendation on the Ethics of Artificial Intelligence*. UNESCO General Conference.
- Veale, M. and Borgesius, F.Z. (2021) Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), pp. 97-112.
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson,

R., Schultz, J. and Schwartz, O. (2018) AI Now Report 2018. AI Now Institute, New York University.

Yeung, K. (2020) Recommendation of the Council on Artificial Intelligence (OECD). *International Legal Materials*, 59(1), pp. 27-34.

Declarations

Funding

This research was supported by the French National Research Agency (ANR) under grant ANR-21-CE26-0022 (GovAI) and a Marie Skłodowska-Curie Individual Fellowship under Horizon 2020 grant 101028614 (ETHICAI). The funder had no role in study design, data collection, analysis, or manuscript preparation.

Conflict of Interest

The author declares no conflicts of interest.

Data Availability Statement

The complete governance framework corpus (84 documents with source URLs), scoring rubric, individual dimension scores, implementation gap analysis, and predictor regression data are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19614596. All analysed frameworks are publicly available from their issuing organisations. Released under CC BY 4.0.

Ethical Approval

This study involved exclusively analysis of publicly available governance documents and did not involve human subjects, surveys, or personal data. No institutional ethical approval was required.

Appendix A

AI Governance Implementation Checklist

This appendix provides a practical checklist for organisations implementing AI governance, derived from best practices across the 84 frameworks analysed.

Fairness and Non-Discrimination

F1. Define fairness metrics appropriate to application context (demographic parity, equal opportunity, or individual fairness)

F2. Conduct pre-deployment fairness audit on protected attributes using disaggregated evaluation

F3. Establish bias monitoring in production with automated alerts for metric drift

F4. Document fairness-accuracy trade-off decisions and rationale

Transparency and Explainability

T1. Maintain model card documentation (intended use, limitations, performance across groups)

T2. Conduct algorithmic impact assessment before deployment of high-risk systems

T3. Provide individual explanations for consequential automated decisions (credit, employment, judicial)

T4. Publish aggregate system performance statistics for public scrutiny

Accountability and Oversight

A1. Designate responsible person/team for each deployed AI system

A2. Establish audit trail capturing model versions, data changes, and decision logs

A3. Implement human oversight mechanisms proportional to decision impact

A4. Create accessible redress mechanism for individuals affected by AI decisions

Enforcement and Compliance

E1. Identify applicable regulatory requirements by jurisdiction and sector

E2. Schedule regular third-party governance audits (annual minimum for high-risk systems)

E3. Report governance incidents (bias findings, safety failures) to oversight body

E4. Maintain compliance documentation suitable for regulatory inspection