AI Ethics Governance Depth: A Cross-National Analysis

Measuring Normative Commitments Across 2,100+ Policies

Lucas Sempé

February 11, 2026

Table of contents

1 AI Ethics Governance Depth: A Cross-National Analysis **3**

Preface **4**
1.1 Key Findings 4
1.2 Methodology 4

2 Introduction **5**
2.1 Measuring Normative Commitments in AI Governance 5

3 Literature Review **6**
3.1 Theoretical Foundations 6

4 Data and Methods **10**
4.1 Shared Methodology 10

5 Ethics Landscape **12**
5.1 The Global Landscape of AI Ethics Governance 12

6 Ethics Determinants **21**
6.1 GDP and Ethics Governance 21

7 Ethics Inequality & Clusters **27**
7.1 Ethics Inequality and Governance Profiles 27

8 Ethics Dynamics **34**
8.1 Convergence, Diffusion, and the Ethics Frontier 34

9 Robustness Checks **44**
9.1 How Robust Are Ethics Findings? 44

10 Discussion **47**
10.1 Implications for Ethics Governance 47

11 Conclusion **50**
11.1 Toward Operationalised Ethics Governance 50

Appendices **52**

1 AI Ethics Governance Depth: A Cross-National Analysis

Measuring Normative Commitments Across 2,100+ Policies

Preface

This study presents the first systematic global assessment of AI ethics governance—measuring the depth and specificity of normative commitments in AI policies worldwide.

Drawing on 2,100+ policies across 70+ jurisdictions, we score each policy on five ethics dimensions: framework depth, rights protection, participatory governance, operationalization, and inclusion. The analysis reveals a landscape where ethical commitments vary more within income groups than between them.

1.1 Key Findings

- *Ethics convergence*: Gaps narrowing over time as developing countries strengthen ethical frameworks
- *Horizontal diffusion*: Regional policy learning dominates over North-South transfer
- *Within-group inequality*: 99% of variation occurs within income groups
- *Documentation confound*: Apparent gaps disappear for well-documented policies

1.2 Methodology

We employ an LLM ensemble (Claude Sonnet 4, GPT-4o, Gemini Flash 2.0) achieving ICC = 0.827 (excellent inter-rater reliability), enabling policy analysis at unprecedented scale.

Citation: Sempé, L. (2026). *AI Ethics Governance Depth: A Cross-National Analysis.* International Initiative for Impact Evaluation (3ie).

Data and Code: github.com/lsempe77/ai-governance-capacity

2 Introduction

2.1 Measuring Normative Commitments in AI Governance

> **Section summary.** How deeply do AI policies embed ethical commitments? This section introduces the research question, examines gaps in ethics governance measurement, and previews our analytical framework.

2.1.1 Beyond Principle Proliferation

The AI ethics landscape shows remarkable normative convergence. Jobin, Ienca, and Vayena (2019) found recurring principles (transparency, fairness, accountability) across 84 guidelines. Fjeld et al. (2020) found similar patterns in 36 frameworks.

But convergence on *principles* doesn't guarantee *governance depth*. Do policies merely list values, or do they specify rights protections, establish participatory mechanisms, operationalize principles, and ensure inclusive engagement?

This study measures ethics governance depth across five dimensions:

1. *E1 Framework Depth*: Specificity of ethical principles and value articulation
2. *E2 Rights Protection*: Safeguards for privacy, fairness, non-discrimination, due process
3. *E3 Participatory Governance*: Public consultation, multi-stakeholder processes, transparency
4. *E4 Operationalization*: Concrete requirements, compliance mechanisms, enforcement
5. *E5 Inclusion*: Representation of marginalized groups, accessibility, equity considerations

Analysis of 2,100+ policies reveals low depth overall (mean 1.39/4.0, with superficial ethical commitments in most policies), narrowing ethics gaps between income groups, extreme within-group inequality (99% of variation within income groups), a documentation confound where the income gap ($d = 0.20$) reverses ($d = -0.09$) for well-documented policies, and horizontal diffusion through peer learning rather than top-down transfer.

The book proceeds as follows: Section 3.1 reviews theoretical foundations; Section 4.1 outlines data and scoring (with full details in the companion *Data, Methods, and Technical Appendices* volume); Section 5.1 maps the global ethics landscape; Section 6.1 examines determinants; Section 7.1 analyses inequality patterns; Section 8.1 traces temporal dynamics and convergence; Section 9.1 presents robustness checks; and Section 10.1 and Section 11.1 discuss implications. The following sections demonstrate that developing countries increasingly match or exceed wealthy-country ethics commitments.

3 Literature Review

3.1 Theoretical Foundations

> **Section summary.** We situate ethics governance measurement within three literatures: AI ethics principles mapping, normative governance frameworks, and the gap between principles and practice.

3.1.1 The AI Ethics Mapping Wave

Jobin, Ienca, and Vayena (2019) analyzed 84 AI guidelines, finding convergence around transparency, fairness, non-maleficence, responsibility, and privacy. Floridi et al. (2018) proposed AI4People, adding beneficence and autonomy. Fjeld et al. (2020) extended mapping to 36 frameworks.

These studies established *what principles appear* but not *how deeply policies operationalize them*. As Hagendorff (2020) observed, the gap between principles and practice remains a pressing challenge for the field. This limitation reflects a methodological constraint: principle-mapping studies typically use binary coding (present/absent) that cannot distinguish superficial mention from substantive operationalization. Moreover, by focusing on high-profile frameworks from major jurisdictions, these studies may overstate global convergence—excluded policies from smaller or developing countries might show different principle priorities.

A second limitation concerns causality: principle convergence might reflect genuine normative consensus, or simply that similar policy communities borrow language without commitment. The temporal pattern matters—did convergence emerge organically from distributed deliberation, or did early frameworks establish templates that subsequent policies copied? Existing studies provide snapshots rather than longitudinal analysis.

From Principles to Governance. Mittelstadt (2019) distinguished *principle-based* from *practice-based* AI ethics. Principles articulate values (fairness, transparency); practice requires translating values into actionable requirements, compliance mechanisms, and enforcement procedures. Most policies remain principle-based.

This distinction is conceptually useful but empirically challenging: what counts as "practice-based" governance? A policy requiring algorithmic impact assessments appears more practice-oriented than one merely endorsing transparency, yet without enforcement mechanisms, compliance monitoring, or sanctions for non-compliance, the requirement may remain aspirational. The principle/practice binary may thus obscure a continuum of operationalization depth.

Rességuier and Rodrigues (2020) proposed measuring ethics "embeddedness"—the degree to which principles become operationalized through concrete requirements. The framework developed here measures this embeddedness across five dimensions. However, Rességuier and Rodrigues developed their framework through qualitative case analysis of 11 European policies, leaving uncertain whether the same embeddedness dimensions apply globally.

3.1.2 Normative Frameworks

Rights-based approaches. Hildebrandt (2019) grounds AI ethics in fundamental rights, arguing that algorithmic systems threaten privacy, equality, and due process. The E2 Rights Protection dimension captures whether policies establish specific safeguards. However, rights traditions vary—European frameworks emphasize data protection, Asian frameworks emphasize collective welfare, African frameworks increasingly emphasize digital sovereignty—raising questions about whether universal rights metrics can capture regional variation. Yeung (2018) shows rights-based governance requires translating abstract rights into technical specifications—a challenging task most policies avoid, which helps account for the low E2 scores documented in Section 5.1.

Participatory governance. Rahwan (2018) argues that AI governance requires democratic participation since algorithmic systems encode societal values. Stilgoe (2020) demonstrates that meaningful participation demands structured processes rather than token consultation. The E3 Participatory Governance dimension measures whether policies establish multi-stakeholder processes, public consultation mechanisms, and transparency requirements. Yet measuring participation structures differs from measuring participation quality—who participates, whose input influences decisions, whether consultations genuinely shape policy.

Operationalisation challenges. Selbst et al. (2019) documents "fairness gerrymandering"—policies proclaiming commitment to fairness without specifying operational definitions or compliance methods. Whittaker et al. (2018) shows this pattern extends to transparency, accountability, and ethics principles broadly. The E4 Operationalisation dimension distinguishes policies that merely mention principles from those specifying concrete requirements, compliance procedures, and enforcement mechanisms.

3.1.3 The Governance Gap

Multiple studies document gaps between commitment and implementation: Mittelstadt (2019), Hagendorff (2020), and Jobin, Ienca, and Vayena (2019) all identify failures to translate principles into tangible governance. This study scores not what policies state but what institutional mechanisms they establish.

Yet these diagnoses rest on small samples and qualitative methods, leaving uncertain how widespread governance gaps are, whether some jurisdictions or policy types perform better, and whether gaps are narrowing over time. Moreover, the literature risks selection bias: researchers naturally examine high-profile frameworks (EU AI Act, Beijing AI Principles, OECD Guidelines) where rhetoric-reality gaps are most visible, potentially overlooking policies from smaller jurisdictions that may operationalize principles more directly precisely because they face less international scrutiny and

reputational pressure. The systematic measurement approach developed here addresses these limitations by scoring all available policies rather than curated samples.

3.1.4 Ethics and Development

Lee (2018) argues developing countries face ethical "catch-up" challenges. Gwagwa et al. (2020) shows African countries increasingly develop indigenous frameworks rather than importing Western principles. Müller (2021) documents how UNESCO's AI Ethics Recommendation (2021) creates opportunity for convergence — which our temporal analysis examines.

However, the "catch-up" framing assumes normative convergence toward Western ethical frameworks as the development endpoint—an assumption that recent scholarship increasingly contests. Gwagwa and colleagues' documentation of indigenous African frameworks suggests plural pathways rather than linear convergence. Moreover, the development literature focuses on whether developing countries *have* ethics frameworks while neglecting whether high-income countries *implement* them—a gap this study addresses by measuring operationalization depth across all income groups. The findings challenge catch-up narratives: developing countries score only marginally lower on ethics depth (Cohen's d = 0.20), and income explains minimal variance once text quality controls are applied.

3.1.5 Measurement Challenges and Contribution

Existing assessments use binary presence/absence (Fjeld et al. 2020) or qualitative evaluation (Mittelstadt 2019). Neither scales to comprehensive global measurement. Binary coding loses information about operationalization depth—a policy mentioning "transparency" scores identically to one establishing mandatory algorithmic impact assessments, external audits, and public reporting. Qualitative evaluation captures depth but requires human expertise that does not scale beyond small samples, typically covering 10-50 policies.

The LLM-based scoring developed here attempts to bridge this trade-off, enabling assessment across 2,100+ policies while preserving depth through multi-level scoring (0-4 scales capturing absence, mention, description, operationalization, comprehensiveness). However, this approach introduces validity concerns: can language models reliably interpret complex institutional arrangements, distinguish meaningful from symbolic commitments, and maintain consistency across diverse legal and policy traditions? The inter-rater reliability analysis (ICC = 0.827) provides empirical reassurance but cannot fully resolve interpretive challenges inherent in cross-national governance measurement.

Contribution. This study addresses four gaps in the existing literature. **First**, principle-mapping studies establish *what* ethics principles appear in AI frameworks but not *how deeply* they are operationalized—this study scores operationalization depth across five dimensions, distinguishing rhetoric from institutional commitment. **Second**, governance gap diagnoses rest on small qualitative samples—this study systematically measures gaps across 2,100+ policies, enabling distributional analysis and statistical testing of determinants. **Third**, the development literature assumes high-income countries lead in ethics governance sophistication—this study tests that assumption

empirically, finding minimal income effects once policy characteristics are controlled. **Fourth**, existing studies provide snapshots while governance evolution remains unexplored—this study examines temporal dynamics, testing whether ethics depth increases over time and whether convergence is occurring.

The findings challenge several conventional assumptions: ethics governance depth varies more within income groups than between them; policies converge on principle *mention* but diverge on *operationalization*; and recent policies show limited improvement over earlier frameworks, suggesting governance stagnation despite proliferating ethics discourse.

4 Data and Methods

4.1 Shared Methodology

This study analyses 2,216 AI policies from the OECD.AI Policy Observatory, scored by a three-model LLM ensemble (Claude Sonnet 4, GPT-4o, Gemini Flash 2.0) on 10 governance dimensions. The full methodological details — corpus construction, scoring rubrics, inter-rater reliability, and technical validation — are documented in the companion volume:

Book 4: Data, Methods, and Technical Appendices

Key parameters for reference:

Table 4.1: Methodology summary

Parameter	Value
Corpus size	2,216 policies, 70+ jurisdictions, 2017–2025
Document retrieval	94% coverage (2,085 full texts)
Analysis-ready text	1,754 documents (79.2%), 11.4 million words
Scoring models	Claude Sonnet 4, GPT-4o, Gemini Flash 2.0
Inter-rater reliability	ICC(2,1) = 0.827 ("Excellent")
Score agreement	95.4% of scores within 1 point across models

4.1.1 Ethics Scoring Framework

Each policy was scored 0–4 on five **ethics governance** dimensions grounded in AI ethics literature (Jobin, Ienca, and Vayena 2019; Floridi et al. 2018; OECD 2019; UNESCO 2021; European Parliament and Council 2024):

Table 4.2: Ethics scoring dimensions

Code	Dimension	What It Measures
E1	Ethical Framework Depth	Grounding in principles, coherent ethical vision
E2	Rights Protection	Privacy, non-discrimination, human oversight, transparency
E3	Governance Mechanisms	Ethics boards, impact assessments, auditing

Code	Dimension	What It Measures
E4	Operationalisation	Concrete requirements, standards, certification
E5	Inclusion & Participation	Stakeholder processes, marginalised group representation

Alongside the five ethics dimensions, each policy was also scored on five parallel **capacity dimensions** (C1 Clarity & Specificity, C2 Resources & Budget, C3 Authority & Enforcement, C4 Accountability & M&E, C5 Coherence & Coordination). The capacity scores inform cross-domain comparisons within this report and are analysed in depth in Book 1. Composite scores are unweighted means: *Capacity* = mean(C1–C5), *Ethics* = mean(E1–E5). The full scoring rubric, validation protocol, and robustness checks appear in Book 4.

Code, data, and methods: https://github.com/lsempe77/ai-governance-capacity

5 Ethics Landscape

5.1 The Global Landscape of AI Ethics Governance

> **Section summary.** This section mirrors the capacity landscape analysis for the five ethics dimensions. Ethics governance lags behind implementation capacity (mean 0.61 vs. 0.83), and 36.3% of all policies score exactly zero on ethics operationalisation.

Having established the scoring methodology and validated the LLM ensemble, this section maps the global landscape of AI ethics governance. While the capacity analysis examined whether governments can implement their policies, the ethics analysis asks a different question: how deeply do those policies engage with normative commitments? The analysis proceeds from aggregate score distributions through income-group comparisons to regional and country-level patterns, revealing an ethics governance landscape that is even weaker than the capacity landscape—but where the North–South divide proves largely illusory.

5.1.1 Overall Score Distribution

Policies worldwide score substantially weaker on ethics than on implementation capacity. Capacity averages 0.83 across the 2,216-policy corpus; ethics averages just 0.61 — a gap of 0.22 points. Policymakers find it easier to specify budgets, agencies, and coordination procedures than to articulate ethical commitments. The floor effects in Figure 5.1 are more severe than for capacity: many policies contain zero ethical content despite explicitly addressing AI governance.

The right-skewed distribution shows that the modal policy contains almost no ethical content, while a minority demonstrates sophisticated ethics governance. Technical implementation dominates; ethical principles remain aspirational.

The ethics composite score averages 0.61/4.00 (SD = 0.62), substantially lower than the capacity composite (0.83):

Table 5.1: Ethics dimension descriptive statistics

Dimension	Mean	SD	Median
E1 Ethical Framework Depth	0.67	0.75	0.00
E2 Rights Protection	0.55	0.66	0.00
E3 Governance Mechanisms	0.62	0.74	0.00
E4 Operationalisation	0.59	0.73	0.00

Dimension	Mean	SD	Median
E5 Inclusion & Participation	0.65	0.73	0.00
Ethics composite	**0.61**	**0.62**	**0.40**

Table 5.1 shows key patterns. The median score of 0.40 falls well below the 2.00 scale midpoint—more than half of policies demonstrate minimal ethical engagement. This compares unfavorably to capacity's median of 0.60. More strikingly, 36.3% of policies score exactly zero on ethics (vs. 27.6% for capacity). Over one-third of the corpus contains no rights protections, no stakeholder inclusion, and no governance frameworks.

Three explanations may account for such pervasive floor effects: (1) many policies address narrow technical issues where drafters consider ethics irrelevant; (2) policymakers deliberately avoid ethical commitments that constrain development or create enforcement burdens; (3) developing ethical frameworks requires expertise, consultation, and political consensus that many jurisdictions lack.

Dimension-level means show little variation (0.55 to 0.67), unlike capacity where Resources (1.15) and Accountability (0.52) diverge sharply. Ethics governance proves uniformly weak. E2 Rights Protection scores lowest (0.55)—fundamental human rights protections remain absent from many policies. E1 Ethical Framework Depth scores highest (0.67) but still below the midpoint; policies citing ethical principles rarely develop them beyond superficial invocations.

Countries demonstrate greater facility in specifying *how* to implement governance than *what ethical standards* should govern implementation. Without ethical foundations, sophisticated implementation infrastructure may serve technocratic efficiency rather than rights protection.

5.1.2 Income-Group Comparisons

The capacity gap between high-income and developing countries was modest (d = 0.30). For ethics, the gap proves smaller still. While capacity requires budgets and technical expertise that wealth facilitates, ethical governance requires political will and normative clarity—factors unrelated to GDP.

These patterns contradict narratives positioning ethics as a "luxury" achievable only after economic growth. Rights protections, stakeholder inclusion, and ethical operationalization prove no more difficult for lower-income countries than for wealthy ones. Developing countries may even hold advantages: fewer entrenched corporate interests resisting governance constraints.

Table 5.2: Income-group ethics comparison

Metric	Value
HI mean (N = 1,700)	0.62
Developing mean (N = 397)	0.50
Welch's t	3.55
p-value	$< .001$
Cohen's d	**0.20**

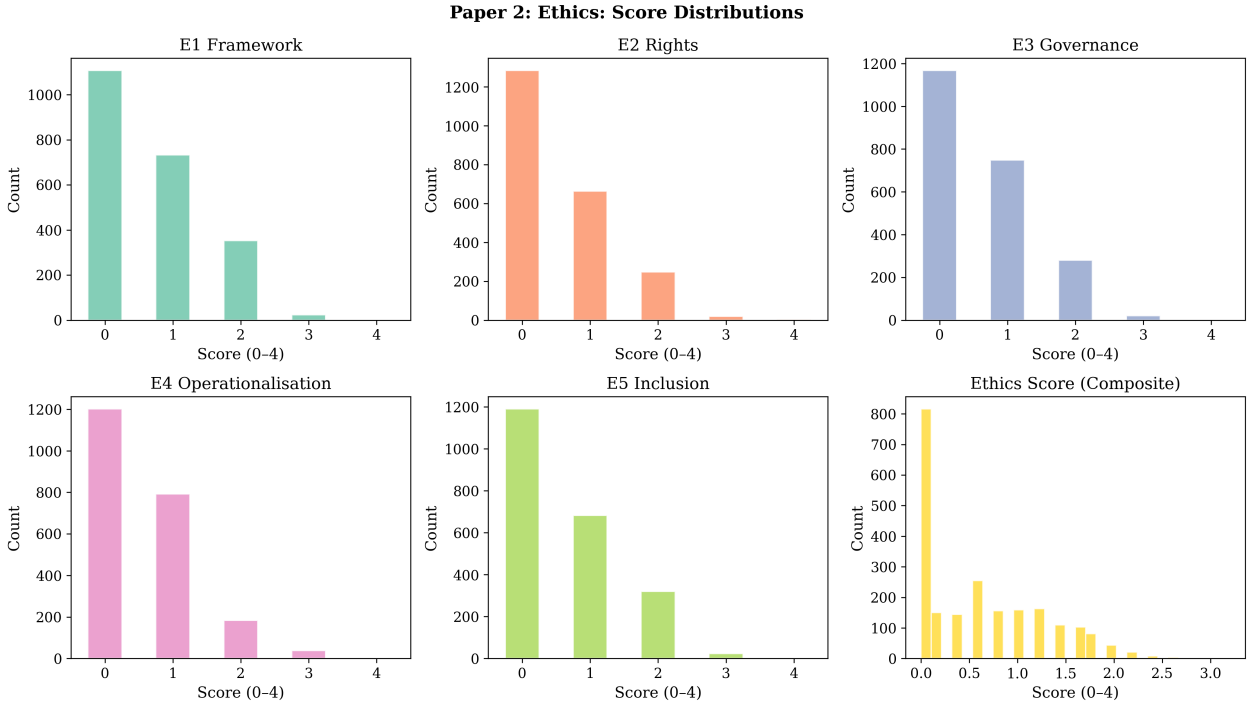

Figure 5.1: Distribution of ethics dimension scores across 2,216 policies. Floor effects are even more pronounced than for capacity.

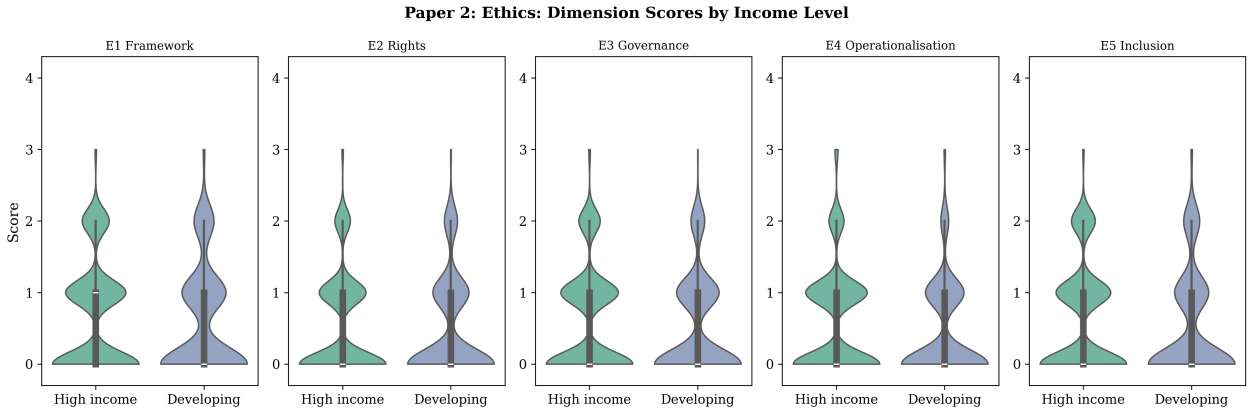

Figure 5.2: Violin plots of ethics scores by income group. The distributions overlap even more than for capacity.

Table 5.2 quantifies what Figure 5.2 shows: the ethics gap is small. Cohen's d = 0.20 falls within the "small effect" threshold (d < 0.30), smaller than capacity's d = 0.30. High-income countries average 0.62 versus 0.50 for developing countries — just 0.12 points on a 4-point scale, about 3% of the range.

This gap vanishes entirely when restricting to good-quality texts (d = -0.09, n.s.; see Section 9.1). The apparent income advantage reflects documentation quality rather than genuine governance differences. Wealthy countries produce longer documents that provide more opportunities to detect ethical content; conditional on text quality, ethical sophistication is unrelated to GDP.

Ethics governance thus requires neither fiscal resources nor technical infrastructure. Countries can build rights-protective, participatory governance frameworks immediately.

Dimension-Level Gaps.

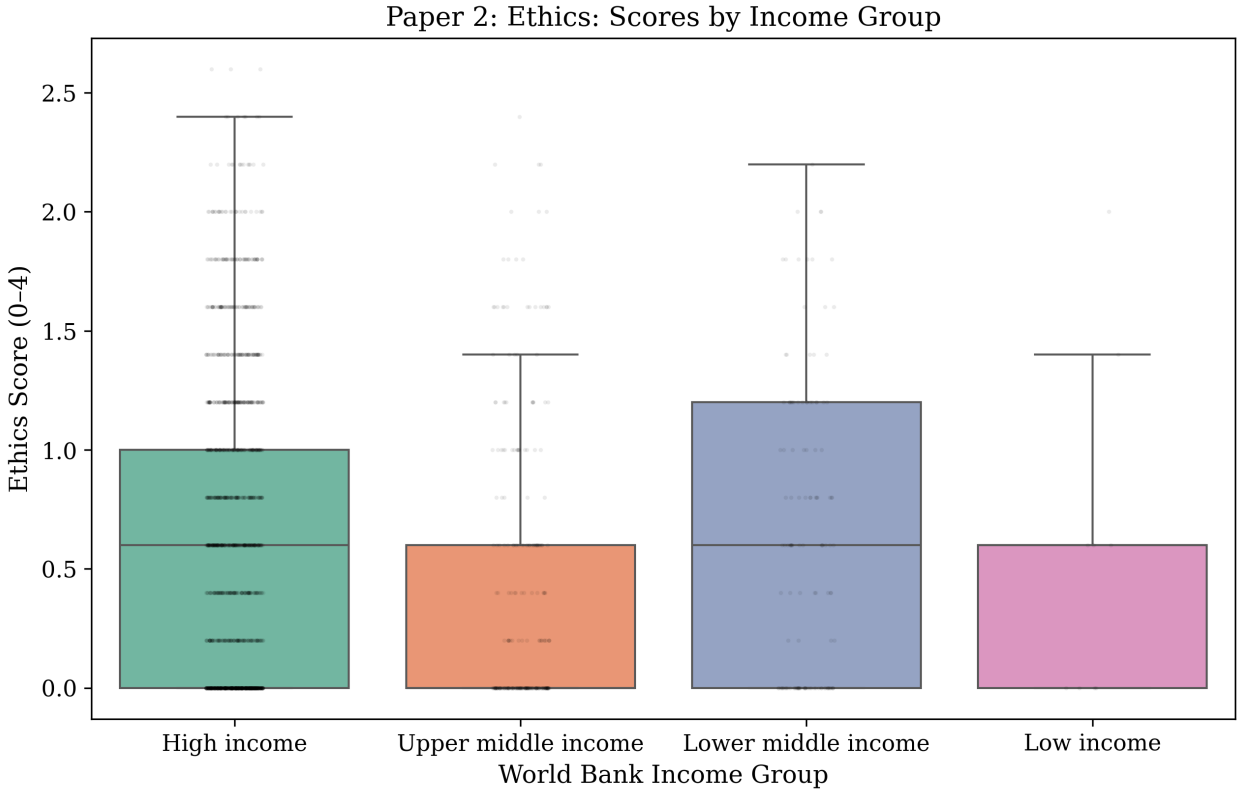

Figure 5.3: Boxplots of ethics scores by income group across all five dimensions.

Table 5.3: Dimension-level income gaps for ethics

Dimension	HI Mean	Dev Mean	Diff	d	p
E1 Framework	0.68	0.58	0.10	0.13	.024
E2 Rights	0.55	0.47	0.08	0.11	.053
E3 Governance	0.63	0.50	0.13	**0.19**	< .001

Dimension	HI Mean	Dev Mean	Diff	d	p
E4 Operationalisation	0.61	0.43	0.17	**0.25**	$< .001$
E5 Inclusion	0.64	0.51	0.13	0.18	.002

Table 5.3 disaggregates the ethics gap by dimension. The largest gap appears on Operationalisation (E4) (d = 0.25, p < .001)—high-income countries translate abstract principles into concrete requirements more successfully. Operationalization demands legal sophistication, auditing capabilities, and administrative infrastructure that wealth can facilitate.

The smallest gap is Rights Protection (E2) (d = 0.11, p = .053). Rights language is universal: developing and wealthy countries alike invoke privacy, non-discrimination, and due process with similar sophistication, reflecting decades of international human rights law. Countries can draw on UDHR Article 12 or ICCPR Article 26 without requiring wealth.

The remaining dimensions—E1 (d = 0.13), E3 (d = 0.19), E5 (d = 0.18)—show small effects. Gaps of 0.08-0.17 points on 4-point scales are substantively negligible.

Income matters for technical operationalization but not for normative commitments. Development interventions should focus on operationalization support (audit tools, impact assessment templates) rather than ethical principles, where developing countries already demonstrate sophistication.

5.1.3 Regional Patterns

If GDP doesn't explain ethics variation, do regional patterns? Shared legal systems, colonial histories, or policy networks might create ethics governance clusters.

Figure 5.4 shows regional variation is less pronounced for ethics than for capacity. Most regions cluster near the global mean of 0.61. North America (0.67) and Europe (0.64) have modest advantages, but smaller than for capacity. Latin America (0.59) approaches the global mean; Sub-Saharan Africa (0.53) lags but not as dramatically as for capacity.

Latin America's near-parity with wealthy regions on ethics governance reflects the region's distinctive political trajectory. Decades of democratisation struggles produced strong civil society organisations, human rights commissions, and constitutional rights traditions that translate naturally into AI governance. Countries like Brazil, Colombia, and Chile approach AI ethics through established frameworks for rights protection and public consultation rather than building from scratch. The Red Iberoamericana de Inteligencia Artificial has also facilitated regional ethics norm-sharing across Spanish- and Portuguese-speaking countries.

East Asia & Pacific presents a paradox: high GDP but modest ethics scores. Countries like Japan, South Korea, and Singapore prioritise competitiveness and innovation in their AI policies, treating ethics as a secondary consideration that might constrain technological development. This "competitiveness-first" framing produces policies heavy on capacity (resources, coordination, authority) but light on rights protections and participatory mechanisms. China's AI governance, while extensive, emphasises state capacity and social management over individual rights protection, further depressing the region's ethics average.

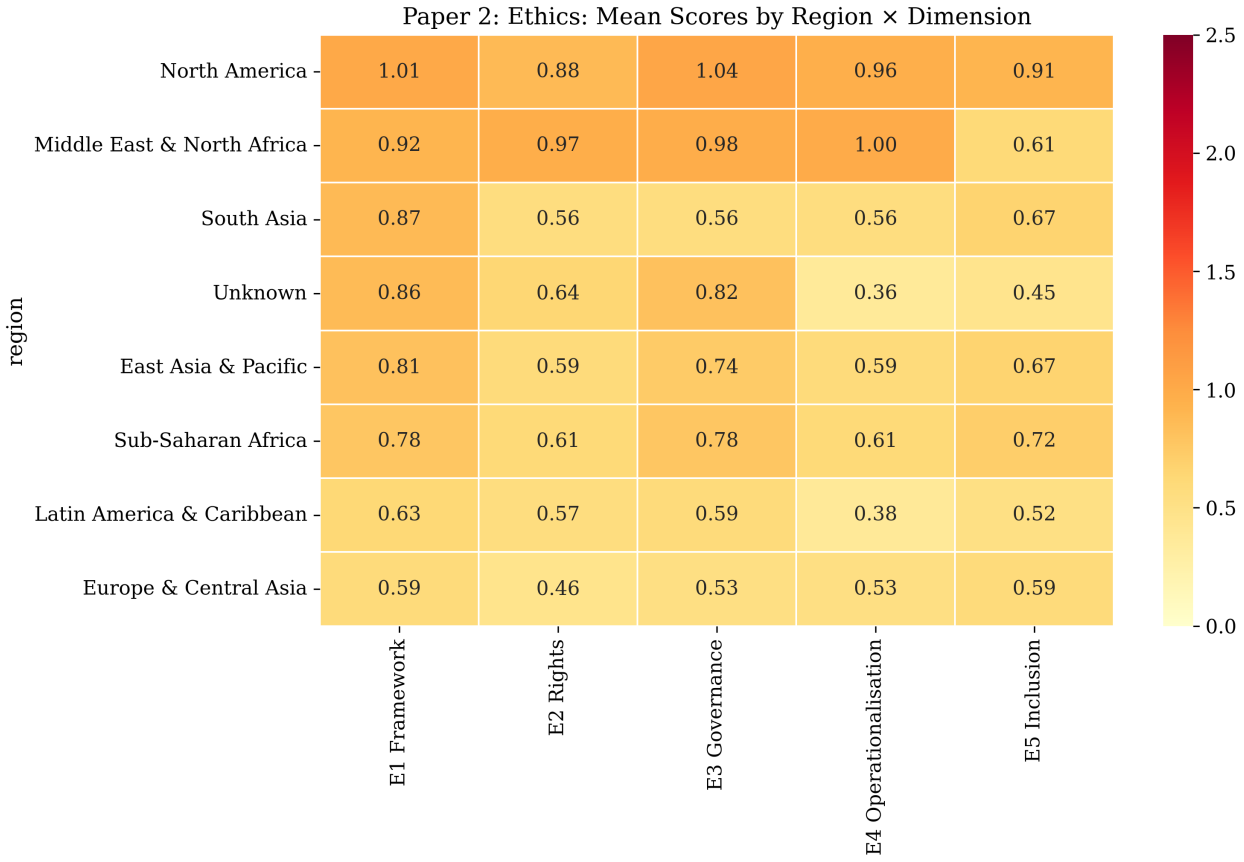

Figure 5.4: Regional heatmap for ethics scores. Regional variation is less pronounced than for capacity.

Sub-Saharan Africa performs relatively better on ethics than on capacity—a notable asymmetry. African countries approaching AI governance often anchor their frameworks in rights-based traditions rooted in the African Charter on Human and Peoples' Rights, the African Union's data governance frameworks, and constitutional protections that emerged from post-colonial state-building. Kenya's Data Protection Act, Rwanda's governance instruments, and South Africa's constitutional rights framework all embed ethics governance within existing rights infrastructure, even where implementation capacity (budgets, agencies, enforcement) remains limited.

MENA and South Asia show the weakest ethics engagement, reflecting both later entry into AI governance and governance cultures that may prioritise state authority and economic development over participatory mechanisms and individual rights protections.

5.1.4 Policy-Type Variation

Do different policy instruments show distinct ethics profiles? Strategies might emphasize ethics while binding regulations focus on compliance.

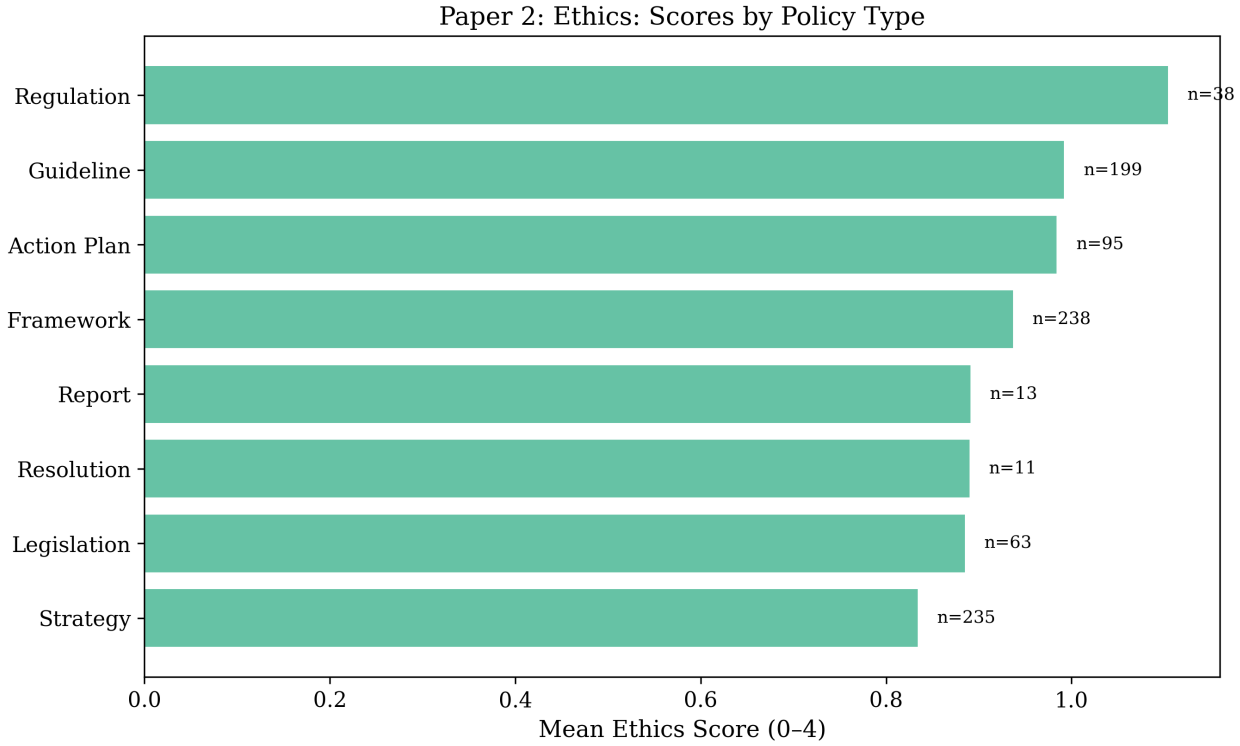

Figure 5.5: Ethics scores by policy type.

Figure 5.5 shows strategies and frameworks score highest (mean 0.75), followed by binding regulations (0.62), with guidance documents trailing (0.55). This differs from capacity, where binding regulation scored highest.

Strategies explicitly aim to articulate normative commitments as their primary purpose. Binding regulations focus on compliance requirements, producing higher capacity but lower ethics scores—they often assume ethical premises implicitly without developing them. Guidance documents address narrow technical issues where ethics may seem tangential.

Comprehensive AI governance requires layered architectures: aspirational strategies for ethical foundations, binding regulations to operationalize principles, and technical guidance for implementation.

5.1.5 Temporal Trends and Correlation Structure

Figure 5.6 previews a notable pattern: overall ethics scores decline modestly from 2017 to 2025, with particularly sharp drops in high-income countries. The mechanisms behind this trend—governance maturation, composition effects, and the shift from aspirational frameworks to technical regulation—are examined in detail in Section 8.1.

Correlation Structure. Do the five ethics dimensions identify separable governance components, or do they collapse into a single "ethics engagement" factor?

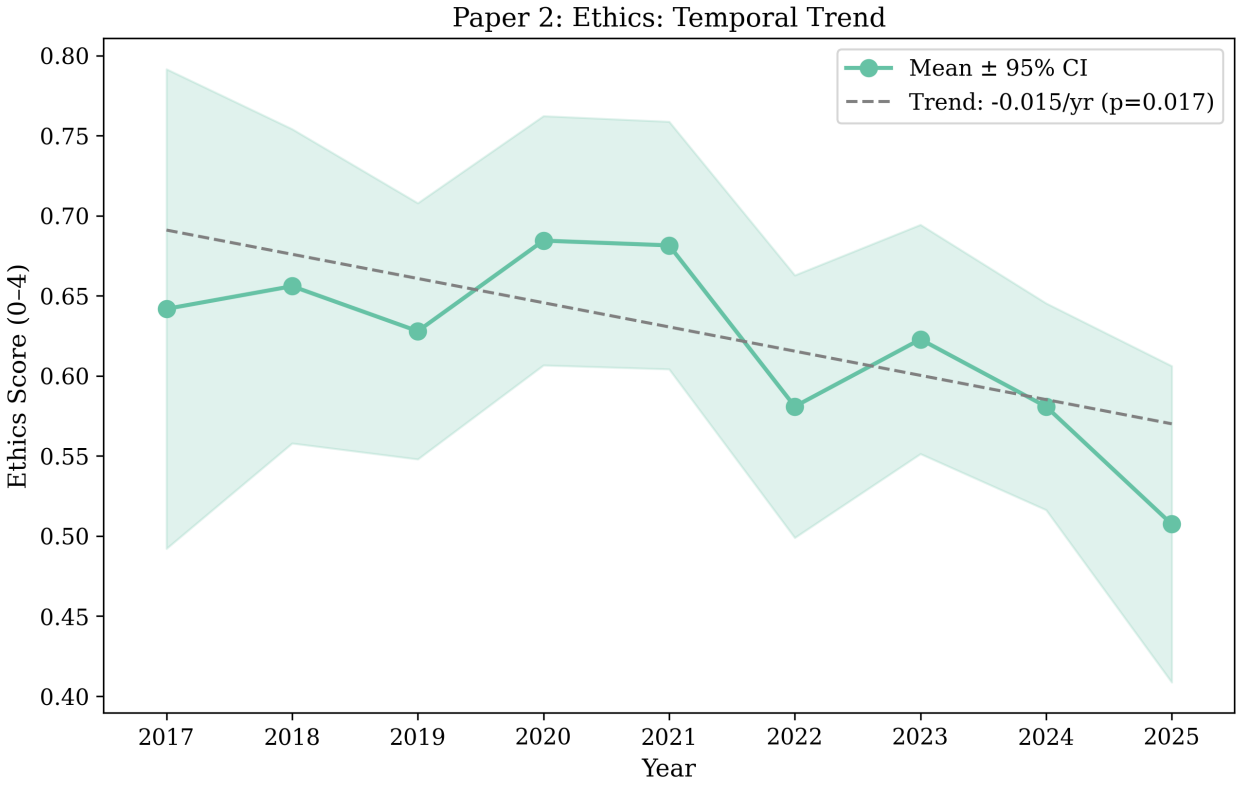

Figure 5.6: Temporal trend in ethics scores (2017–2025).

Figure 5.7 shows moderate-to-strong positive correlations, all exceeding r = 0.40. The strongest: E1 Framework Depth and E4 Operationalisation (r 0.72)—policies with sophisticated frameworks also tend to operationalize them. Frameworks without operationalization produce aspirational documents; operationalization without frameworks produces normatively hollow governance.

Similarly, E3 Governance Mechanisms and E5 Inclusion correlate at r 0.68. Accountability structures and stakeholder inclusion go together: mechanisms without input risk technocratic disconnection; inclusion without mechanisms risks empty consultation.

Despite these correlations, the dimensions remain separable. The PCA analysis confirms that capacity and ethics (r = 0.75) constitute empirically distinct constructs — policies can have sophisticated capacity with minimal ethics, or vice versa.

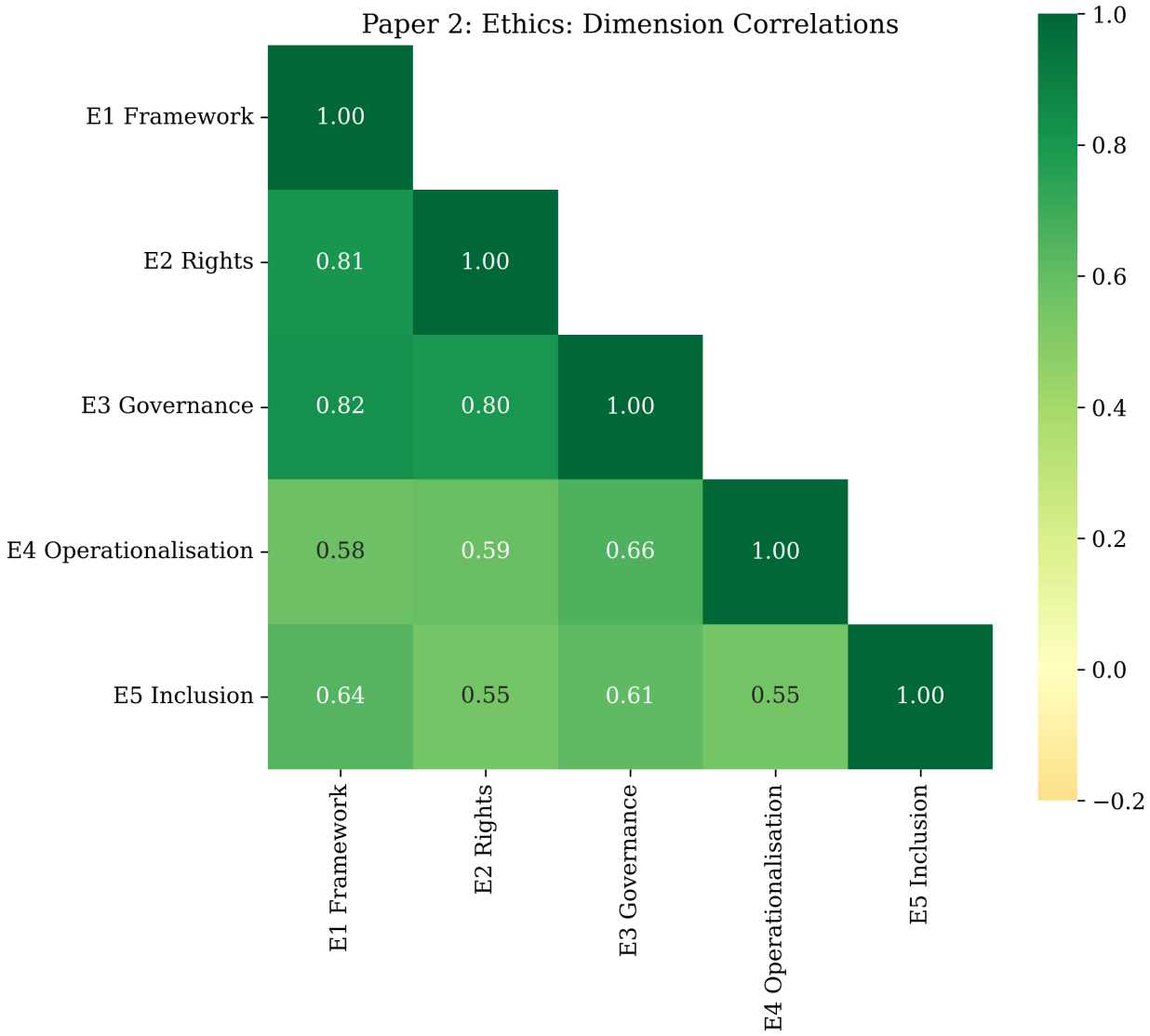

Figure 5.7: Correlation matrix for the five ethics dimensions.

6 Ethics Determinants

6.1 GDP and Ethics Governance

> **Section summary.** This section examines a key asymmetry between capacity and ethics: while GDP has a modest positive effect on capacity, it shows no effect on ethics scores at any quantile. The OLS significance is entirely driven by the extensive margin (whether any policy exists).

The landscape analysis documented low ethics scores overall, with over a third of policies scoring zero. This section asks what drives variation in ethics governance depth. A common assumption treats ethical governance as a luxury—something only wealthy nations can afford after meeting basic institutional needs. The regression analysis that follows tests this assumption through OLS, quantile, and Tobit specifications. The results reveal a striking asymmetry with the capacity domain: whereas GDP exerts a modest effect on implementation capacity, it shows virtually no effect on ethics governance at any point in the distribution.

6.1.1 OLS Regression

Figure 6.1 shows a weak positive association with substantial scatter. The OLS regression:

Table 6.1: OLS regression for ethics (selected coefficients)

Variable	β	SE	t	p
log(GDP pc)	0.061	0.020	3.05	.002
Good text quality	1.014	—	—	$< .001$

Table 6.1 shows = 0.061 (p = .002), apparently supporting the "luxury good" hypothesis. However, as quantile regression reveals below, this significance represents a statistical artifact.

Text quality dominates (= 1.014, p $< .001$): well-documented policies score a full point higher than poorly documented ones. This effect exceeds that of GDP.

The methodological issue: OLS assumes GDP's effect remains constant across the distribution. However, with 36.3% of policies scoring exactly zero, GDP may operate through the *extensive margin* (whether any ethics content exists) rather than the *intensive margin* (sophistication level among policies with ethics content).

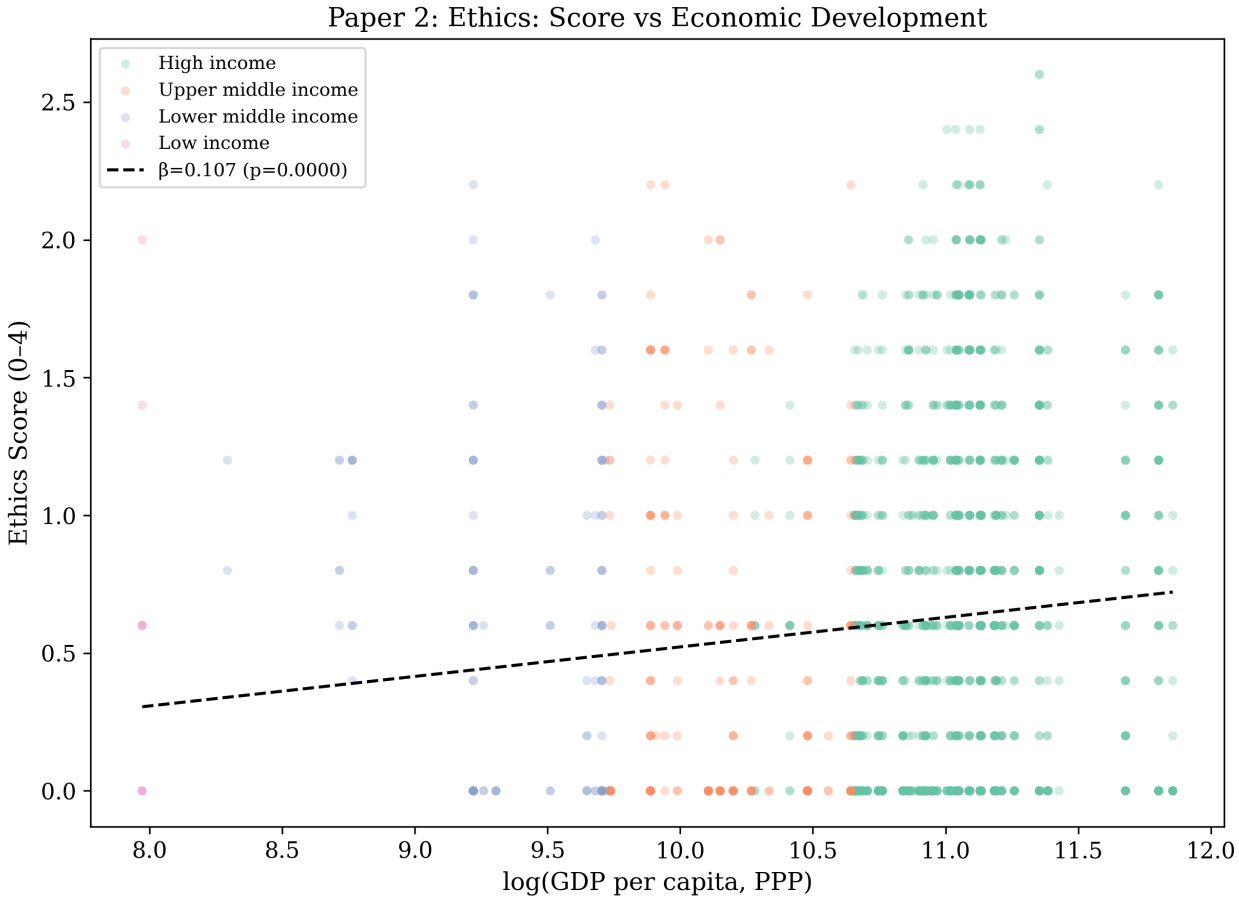

Figure 6.1: Scatter plot of ethics scores against log GDP per capita.

6.1.2 Quantile Regression: The Zero-Effect Finding

Quantile regression estimates GDP's effect at different distribution points. If GDP facilitates ethics throughout, coefficients should be positive at all percentiles. If GDP affects only the extensive margin, coefficients should approach zero at all positive-score quantiles.

Figure 6.2 delivers the central finding: the regression line is flat at 0 across the entire distribution.

Table 6.2: Quantile regression: GDP effect across the ethics distribution

Quantile (τ)	GDP β	SE	p
0.25, 0.50, 0.75	0.000	—	n.s.
OLS (reference)	0.061	0.020	.002

GDP shows no effect at any quantile examined. The OLS significance arises entirely from the extensive margin—wealthier countries produce more policies overall, mechanically increasing the

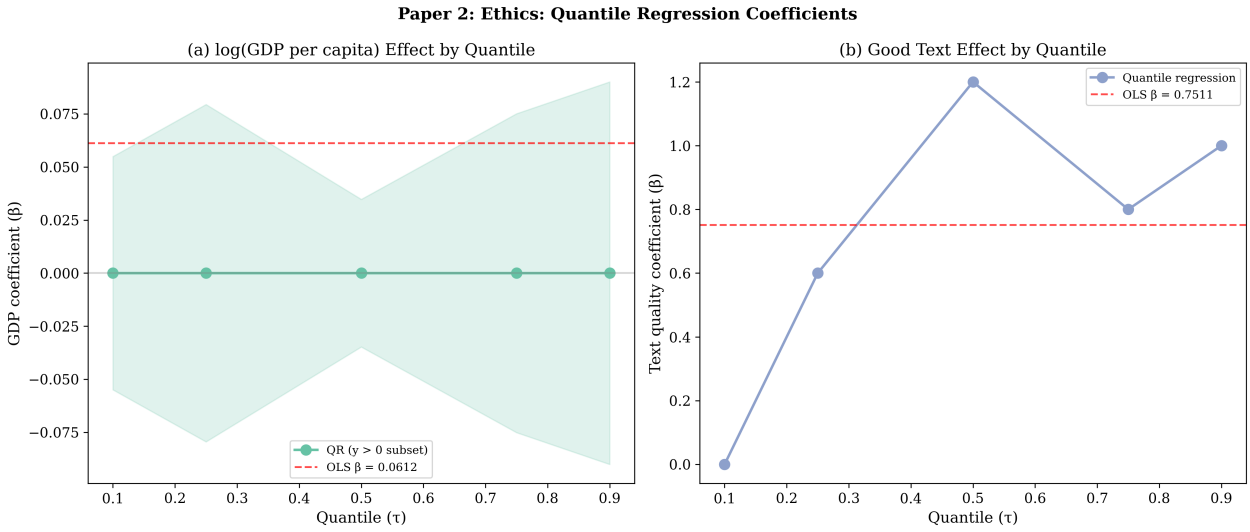

Figure 6.2: Quantile regression coefficients for GDP across the ethics distribution. GDP shows no effect at any quantile—a flat line at $= 0$.

probability that at least one contains ethics content. Conditional on having a non-zero ethics score, GDP contributes nothing.

This contradicts the "luxury good" hypothesis. Countries at any income level produce policies with sophisticated ethics frameworks. Kenya's AI and Data Protection regulations, Brazil's AI Ethics Framework, and Colombia's AI strategy all demonstrate high ethics scores despite modest GDP levels. Conversely, wealthy countries often produce ethics-light policies focusing on technical compliance.

The asymmetry with capacity is interpretable: capacity requires fiscal resources and infrastructure; ethical commitment requires political will and stakeholder engagement—factors unrelated to GDP.

For development interventions, these findings suggest that promoting ethical AI governance need not wait for economic growth. Supporting political processes that enable ethical deliberation, strengthening civil society organizations advocating for rights, and providing operationalization assistance can proceed regardless of income level.

6.1.3 Multilevel and Tobit Models

Quantile regression treated policies as independent. Multilevel modeling accounts for nesting within countries, correcting for dependency that inflates standard errors. The key parameter is the Intraclass Correlation Coefficient (ICC)—the proportion of variation occurring between vs. within countries.

Figure 6.3 shows country random effects. Iceland, Rwanda, and Nigeria consistently produce higher-ethics policies than GDP predicts; Kazakhstan and several wealthy Asian countries underperform.

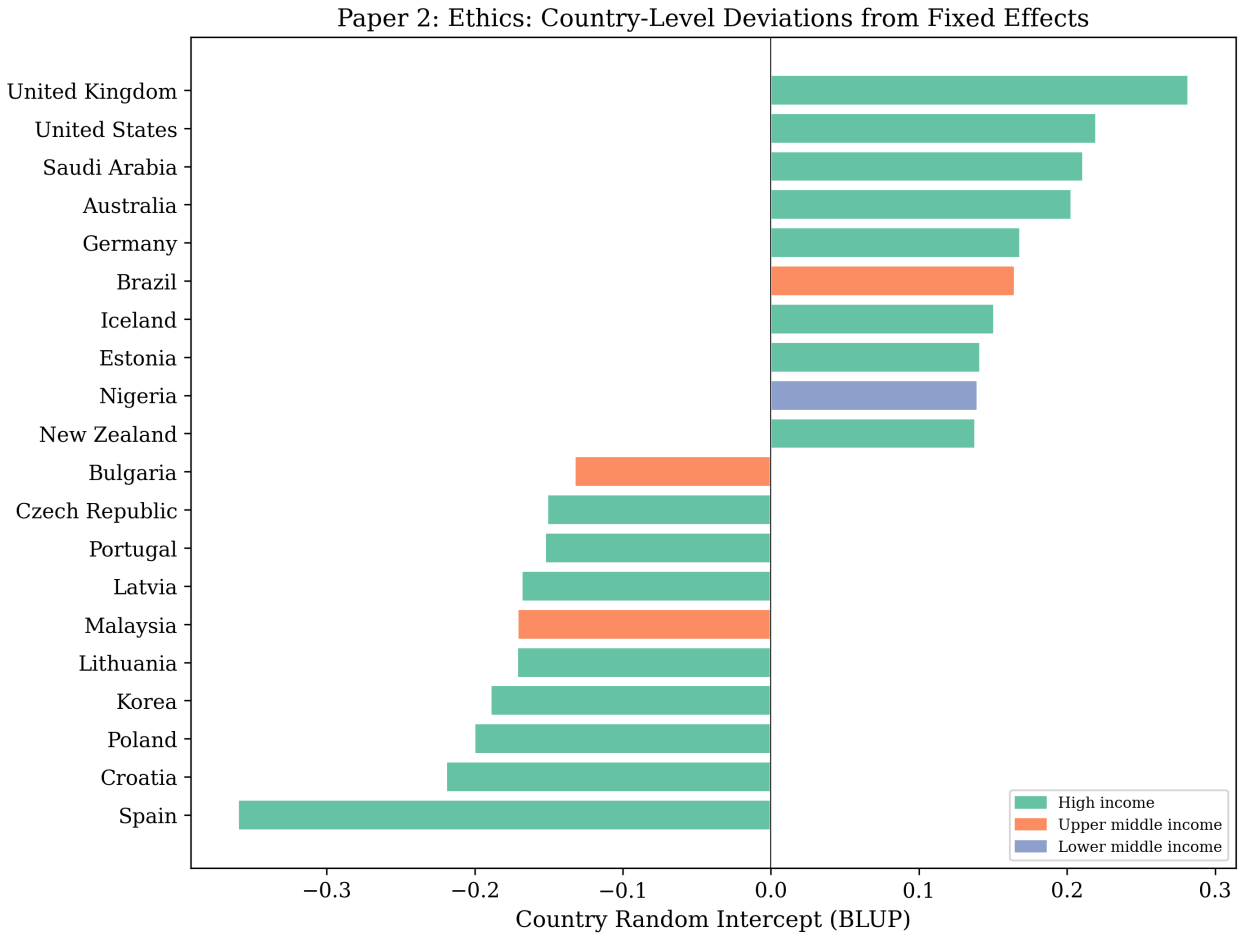

Figure 6.3: Country random effects from the multilevel ethics model.

Table 6.3: Multilevel model comparison for ethics

Metric	OLS	Mixed
GDP β	0.061	0.029
GDP p	.002	.38
Country ICC	—	**0.125**

Two findings from Table 6.3: (1) The GDP coefficient collapses from = 0.061 (p = .002) to = 0.029 (p = .38)—a 52% reduction and complete loss of significance. The multilevel model properly attributes within-country correlation to random effects rather than spuriously to GDP. (2) Country ICC = 0.125: 12.5% of variation is between countries; 87.5% is within countries. Policy-specific factors—document purpose, authoring agency, consultation processes—matter far more than national characteristics.

The random effects identify consistent over-performers (Iceland, Rwanda, Nigeria) — countries where political commitment to ethics exceeds GDP predictions.

Tobit Regression. With 36.3% zero-score rates (exceeding capacity's 27.6%), OLS and quantile regression may be biased. Zero-scoring policies represent complete absence of ethics, not "minimal ethics." Tobit regression models this censoring.

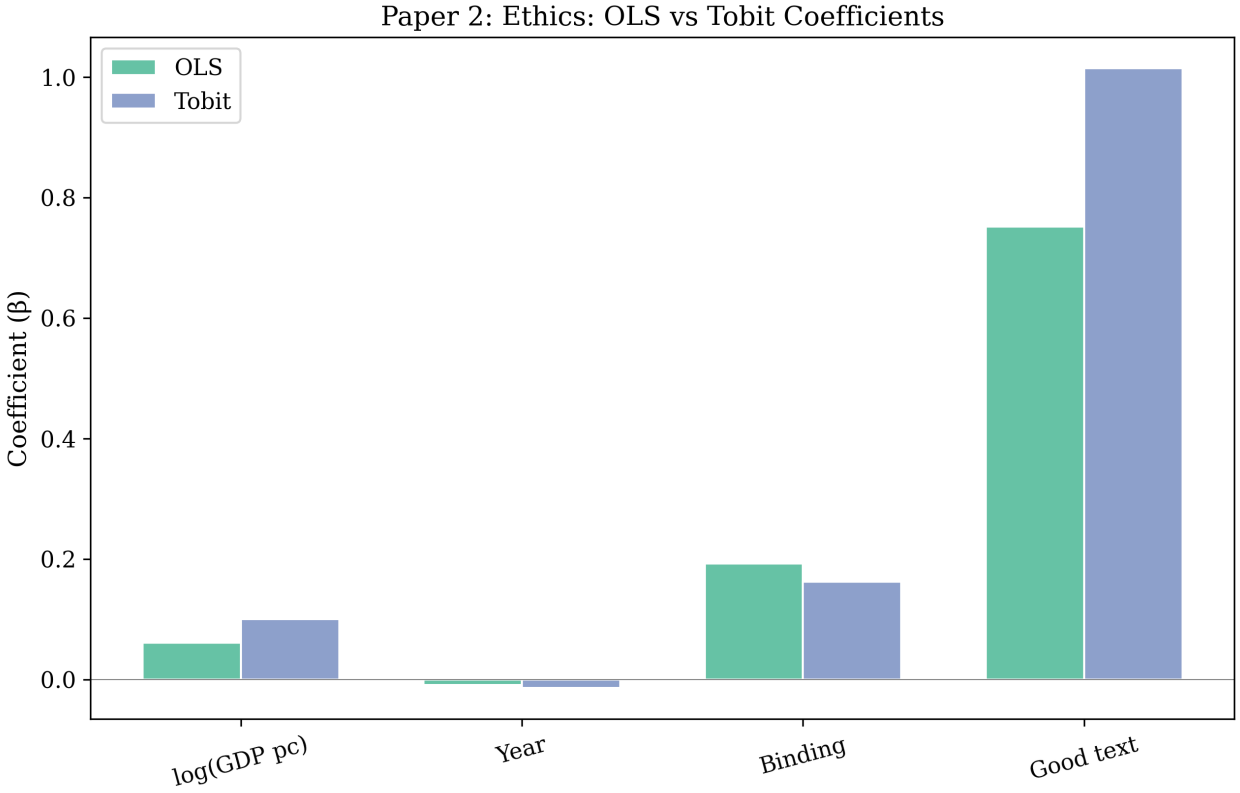

Figure 6.4: Comparison of OLS and Tobit coefficients for the ethics model.

Table 6.4: Tobit model for ethics

Variable	OLS β	Tobit β
log(GDP pc)	0.061	0.100
Good text quality	—	1.014
σ	—	0.700
P(uncensored at mean)	—	0.725
Floor: score = 0	36.3%	—

Censoring correction increases the GDP coefficient from = 0.061 (OLS) to = 0.100 (Tobit). But interpreted alongside the quantile finding: this effect operates entirely through the extensive margin (reducing probability of zeros) rather than improving positive scores. Wealthy countries are more likely to produce *some* ethics content, but conditional on having content, GDP contributes nothing to quality.

Synthesis. GDP does not predict ethics governance quality. OLS shows nominal significance (=

0.061, p = .002), but this proves artifactual: quantile regression reveals no effect at any quantile, and multilevel modeling eliminates significance entirely (= 0.029, p = .38). The Tobit model confirms that GDP operates only through the extensive margin—wealthier countries are more likely to produce *some* ethics content, but conditional on having content, GDP contributes nothing to quality.

The asymmetry with capacity is interpretable: implementation requires resources that wealth provides, whereas ethical governance requires political will, democratic traditions, and civil society strength—factors independent of GDP. Kenya, Colombia, Brazil, and Rwanda demonstrate sophisticated ethics frameworks despite modest GDP. Kazakhstan's resource wealth fails to generate ethics governance. Promoting ethical AI governance need not wait for economic growth.

7 Ethics Inequality & Clusters

7.1 Ethics Inequality and Governance Profiles

Section summary. Inequality decomposition for ethics mirrors the capacity findings: 99.5% of variation is within income groups. Portfolio breadth is near-universal, but E2 Rights and E5 Inclusion show the largest coverage gaps.

The determinants analysis showed that GDP fails to predict ethics governance depth. This section examines where variation in ethics scores actually resides. If the governance gap is not primarily between income groups, then the conventional framing of a global ethics divide requires revision. Using the same inequality decomposition methods applied to capacity—Gini coefficients, Lorenz curves, and Theil indices—this section demonstrates that 99.5% of ethics variation occurs within income groups, an even more extreme concentration than the 98.8% observed for capacity.

7.1.1 Inequality Decomposition

Capacity analysis demonstrated 98.8% of governance variation occurring within income groups. The ethics domain exhibits a similar pattern.

Gini coefficients, Lorenz curves, and Theil decomposition are employed to partition ethics inequality into between-group and within-group components.

Figure 7.1 shows the developing-country curve bows further from the equality line, indicating greater heterogeneity.

Table 7.1: Gini coefficients for ethics scores

Metric	Value
Gini (all countries)	0.569
Gini (HI only)	0.553
Gini (Developing)	0.638
Gini (country means)	0.273

Table 7.1 shows ethics inequality exceeds capacity inequality (0.569 vs. 0.518 Gini), reflecting the severe floor effect (36.3% zeros). Ethical engagement proves polarized: countries either commit substantially or largely ignore ethical dimensions.

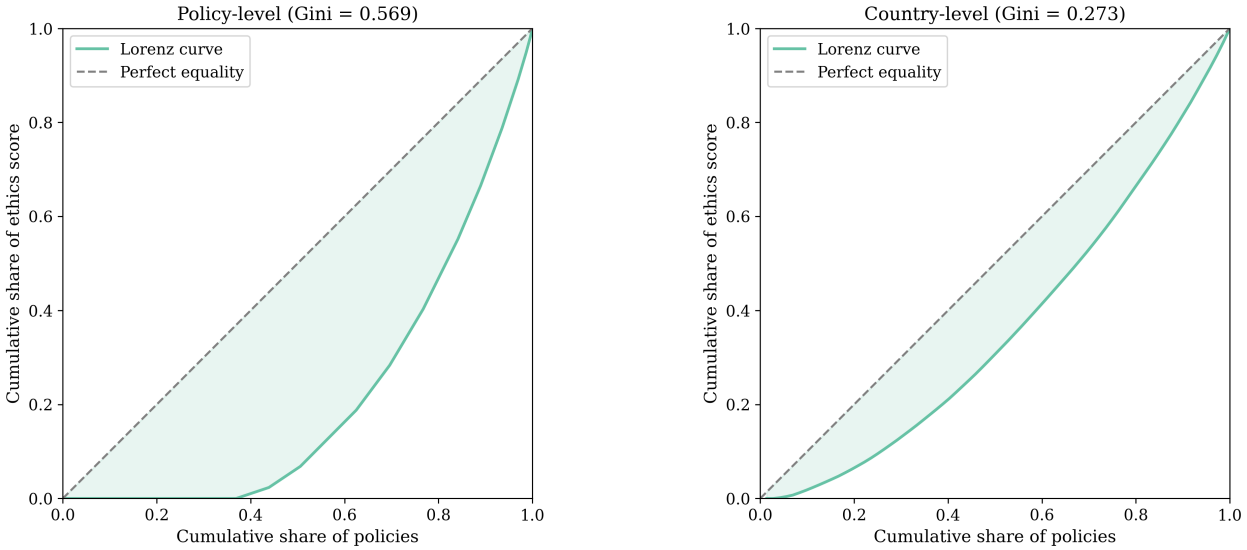

Figure 7.1: Lorenz curves for ethics scores. Developing countries show higher within-group inequality than HI countries.

A key asymmetry emerges: developing countries show higher inequality (0.638) than high-income countries (0.553). Developing countries span the full spectrum—Kenya, Colombia, Brazil, and Rwanda match or exceed many high-income countries, while others produce near-zero policies. High-income countries converge on moderate scores with fewer extremes.

The country-means Gini (0.273) is lower because averaging smooths policy-level variation. Some countries consistently outperform others, reflecting political and institutional factors rather than wealth.

Theil Decomposition. Theil's index enables exact additive decomposition of inequality into between-group and within-group components.

Figure 7.2 shows the between-group slice is barely visible.

Table 7.2: Theil decomposition for ethics

Component	Share
Between income groups	0.5%
Within income groups	99.5%

Only 0.5% of ethics inequality is between income groups; 99.5% is within. This is even more extreme than capacity's 1.2%. Knowing whether a policy comes from a high-income or developing country tells you almost nothing about its ethics score.

For development policy, these findings suggest economic growth alone would do little to improve ethics governance. A country crossing the high-income threshold gains 0.12 points (3% of the

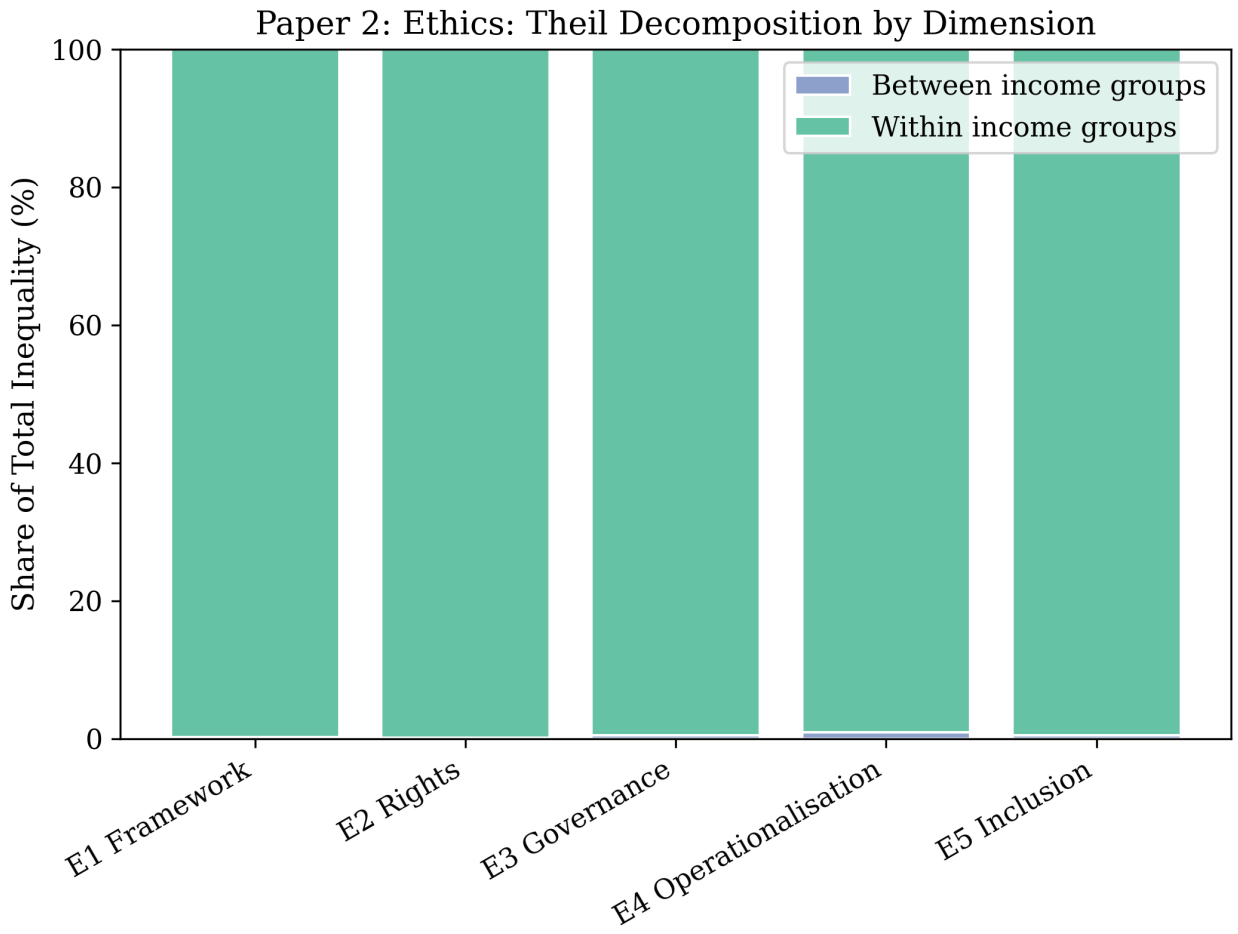

Figure 7.2: Theil decomposition: 99.5% of ethics inequality is within income groups.

scale) but still faces the same within-group variation. Targeting factors that matter—political prioritization, civil society strength, stakeholder inclusion—appears more promising than assuming growth generates ethics.

7.1.2 Policy Portfolios and Governance Typologies

Countries produce portfolios of policies. *Breadth* measures how many of the five ethics dimensions receive non-zero scores in at least one policy. A 5/5 country addresses all dimensions; 3/5 has gaps on two.

Table 7.3: Ethics portfolio breadth

Metric	HI	Developing	p
Mean breadth (out of 5)	5.00	4.36	.054

Metric	HI	Developing	p
Countries with 5/5 coverage	94%	—	—
Least covered dimensions	E2 Rights (94.1%)	E5 Inclusion (94.1%)	—

Table 7.3 shows the gap is marginally significant (p = .054). High-income countries achieve near-universal coverage (94% at 5/5); developing countries average 4.36/5 with modest gaps. E2 Rights and E5 Inclusion are most commonly absent—a pattern that deserves scrutiny.

The dimensions most likely to be missing are precisely those that most directly protect marginalised populations from algorithmic harms. This is unlikely to be coincidental. Rights protection (E2) requires governments to constrain their own authority and that of powerful private actors, creating enforcement obligations and potential litigation exposure. Inclusion (E5) demands governments actively engage groups—ethnic minorities, disabled persons, rural populations, women—that may lack political power to demand representation. Both dimensions impose costs on powerful actors and benefit diffuse, politically weak constituencies. By contrast, E1 Framework Depth (articulating principles) and E3 Governance Mechanisms (establishing oversight bodies) are politically less costly: they signal commitment without necessarily constraining behaviour. The portfolio gap pattern thus reflects a political economy of ethics governance, where countries adopt the dimensions that are cheapest to signal and avoid those that require genuine redistribution of power.

Most countries engaging ethics do so comprehensively rather than cherry-picking, contrary to narratives suggesting superficial ethics adoption for international legitimacy.

Cluster Analysis. Cluster analysis identifies distinct ethics profiles — typologies defined by dimensional strengths and weaknesses.

The two-cluster solution (silhouette = 0.42) mirrors capacity:

Cluster 1 ("Ethics-Light"): ~65% of policies. Near-zero scores across all dimensions. These policies address AI through purely technical lenses—procurement guidelines, technical standards—without engaging ethics.

Cluster 2 ("Ethics-Engaged"): ~35% of policies. Moderate scores emphasizing E1 Framework Depth and E3 Governance Mechanisms. These explicitly articulate principles and establish accountability structures. National strategies and comprehensive frameworks fall here.

Both clusters contain policies from both income groups. Developing countries produce both Ethics-Light and Ethics-Engaged policies. High-income countries span both clusters. Income does not determine governance typology.

The binary characterizes global AI governance: countries either commit to comprehensive normative engagement or focus narrowly on technical implementation. Effective governance requires both types — ethics frameworks establishing foundations, plus technical policies implementing them.

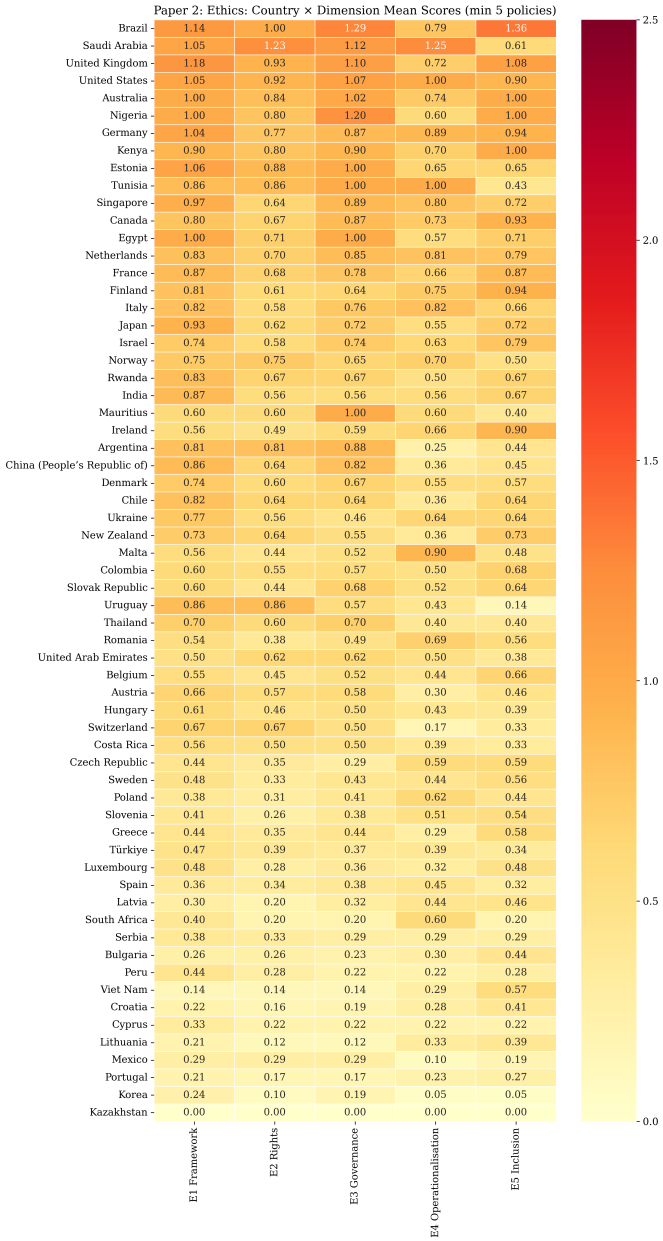

Figure 7.3: Policy portfolio coverage for ethics dimensions.

Figure 7.4: Ethics portfolio gap analysis.

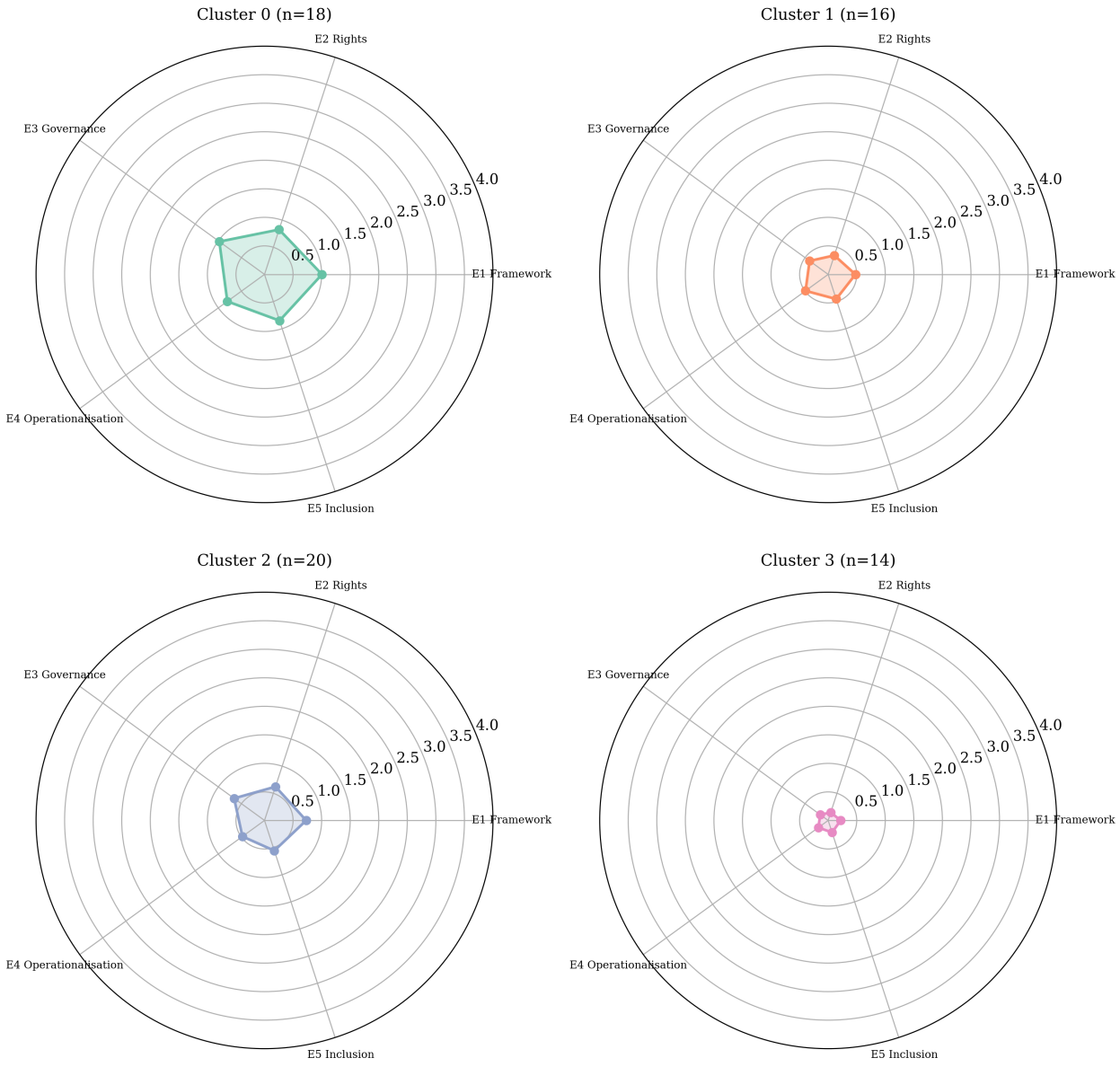

Figure 7.5: Cluster profiles for ethics dimensions. Two-cluster solution mirrors the capacity pattern.

8 Ethics Dynamics

8.1 Convergence, Diffusion, and the Ethics Frontier

> **Section summary.** In contrast to capacity, ethics scores are converging across income groups—though through an unexpected mechanism: high-income countries are declining rather than developing countries improving. The efficiency frontier for ethics features several African countries exceeding GDP-based predictions.

The cross-sectional analyses established that ethics scores are low, GDP is irrelevant as a predictor, and nearly all variation lies within income groups. This section introduces a temporal dimension: are these patterns stable, or are they shifting? The capacity dynamics analysis found no convergence—income-group gaps remained flat from 2017 to 2025. Ethics tells a different story. Gaps are narrowing, but through an unexpected mechanism that challenges the standard narrative of developing countries "catching up" to wealthy-nation standards.

8.1.1 The Convergence Finding

Capacity demonstrated stable gaps from 2017-2025. Ethics diverges: gaps are narrowing.

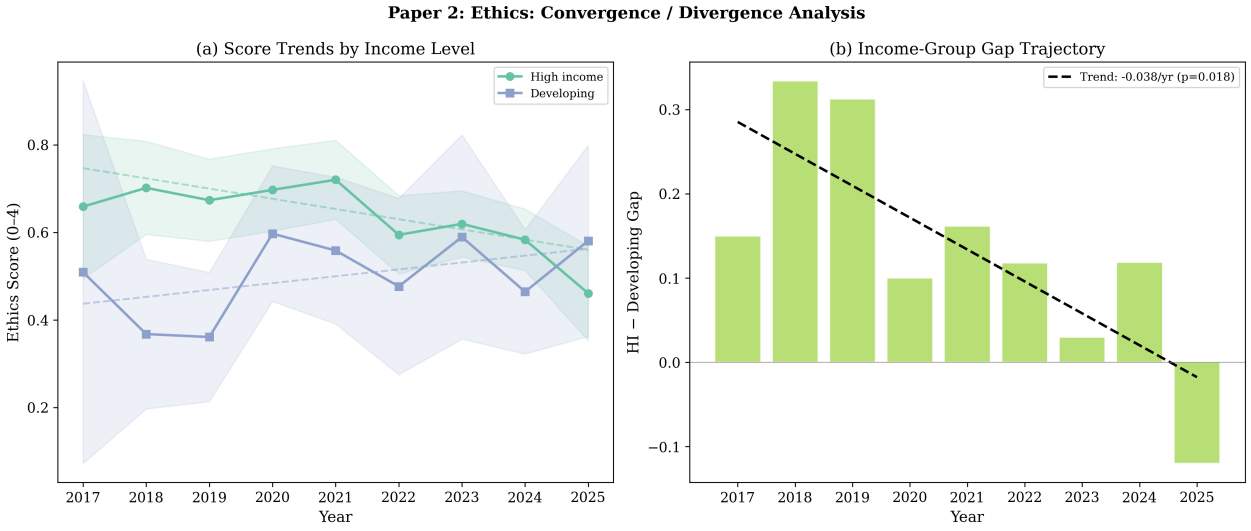

Figure 8.1: Ethics convergence trends by income group. The gap is narrowing, driven by HI countries declining.

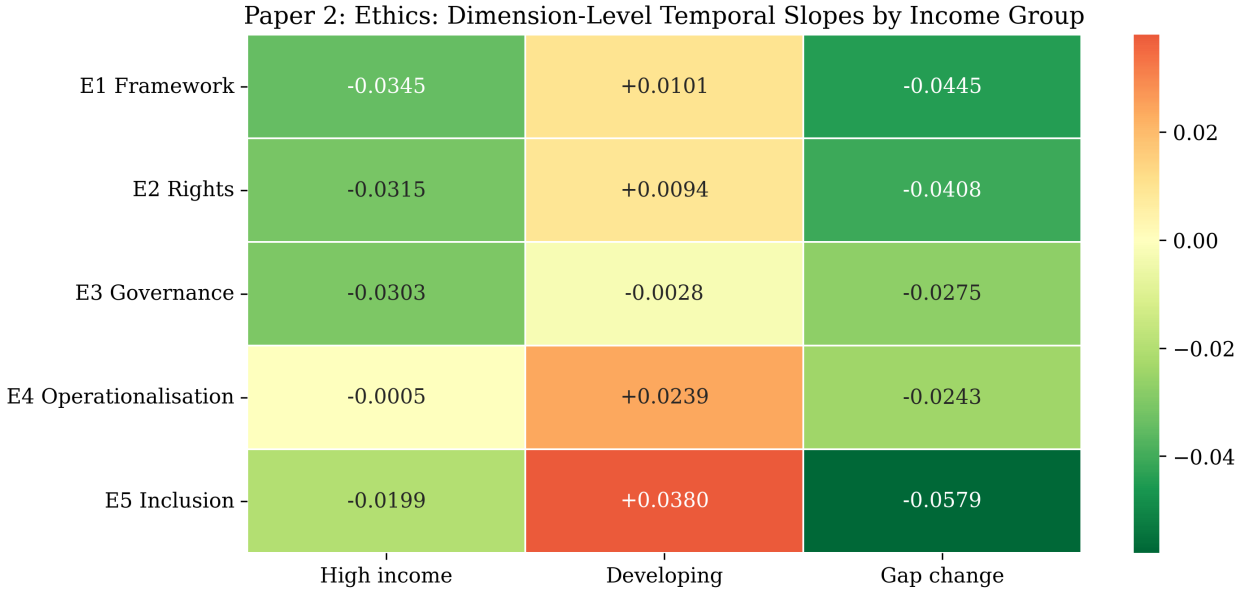

Figure 8.2: Dimension-level convergence for ethics.

Figure 8.1 shows high-income countries declining while developing countries remain flat. Figure 8.2 shows the decline affects multiple dimensions.

Table 8.1: Convergence test for ethics scores

Metric	Value
Income × Year interaction	$\beta = -0.031$, $p = .015$*
HI temporal slope	-0.023/yr ($p = .001$)
Developing temporal slope	$+0.016$/yr (n.s.)
Gap trend	Narrowing (-0.038/yr, $p = .018$)

Ethics gaps narrow at -0.031 points per year (p = .015). The mechanism proves unexpected: high-income countries decline at -0.023/yr (p = .001), while developing countries show modest, non-significant improvement ($+0.016$/yr). The gap narrows because wealthy countries retreat rather than developing countries advancing.

Several explanations may account for high-income country declines:

1. *Regulation replacing aspiration:* Early policies (2017-2020) were deliberative frameworks—Canada's Declaration, France's Villani Report. Later policies (EU AI Act) focus on compliance, assuming ethics as background. This represents maturation rather than regression.
2. *Ethics fatigue:* Initial enthusiasm waned as AI governance became routine. Newer policies address technical issues where ethics may seem tangential.
3. *Composition effects:* Later policies are narrower (AI procurement in healthcare) rather than comprehensive frameworks.

4. *Industry pressure:* Some countries de-emphasize ethics amid arguments that requirements stifle competitiveness.

Developing countries show modest improvement (+0.016/yr), though below significance thresholds — indicating heterogeneity consistent with 99.5% within-group variation. If trends continue, the gap closes entirely by 2030. Ethics diffuses more readily than capacity, requiring political will rather than fiscal resources.

8.1.2 Temporal Trends

Not all dimensions decline equally; specific dimensions drive the observed trend.

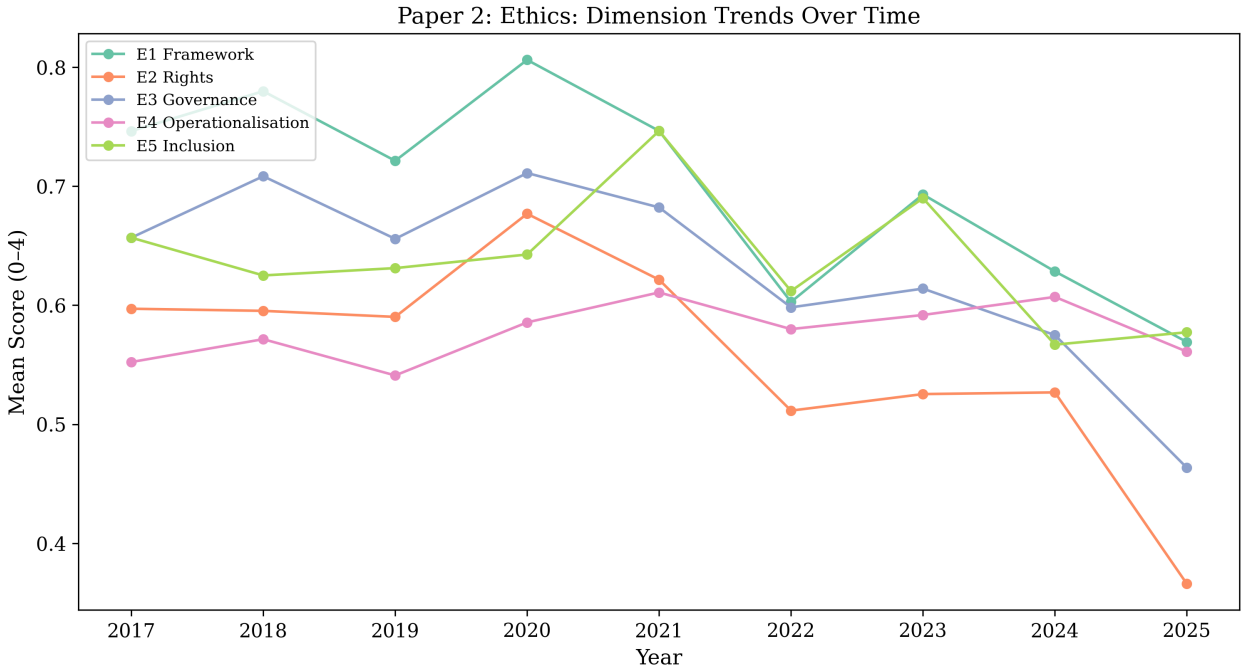

Figure 8.3: Ethics dimension scores over time.

Figure 8.3 shows E1 Framework Depth and E3 Governance Mechanisms decline sharpest, while E2 Rights Protection and E5 Inclusion remain stable. This supports the "regulation replacing aspiration" interpretation: newer policies assume rights as established principles while focusing on implementation details rather than comprehensive frameworks.

Figure 8.4 confirms convergence visually: high-income trajectories trend down, developing countries stay flat, and the gap narrows.

8.1.3 Diffusion and the Ethics Frontier

Capacity showed 98% horizontal diffusion (peer learning) rather than vertical (rich-to-poor). Ethics exhibits a similar pattern.

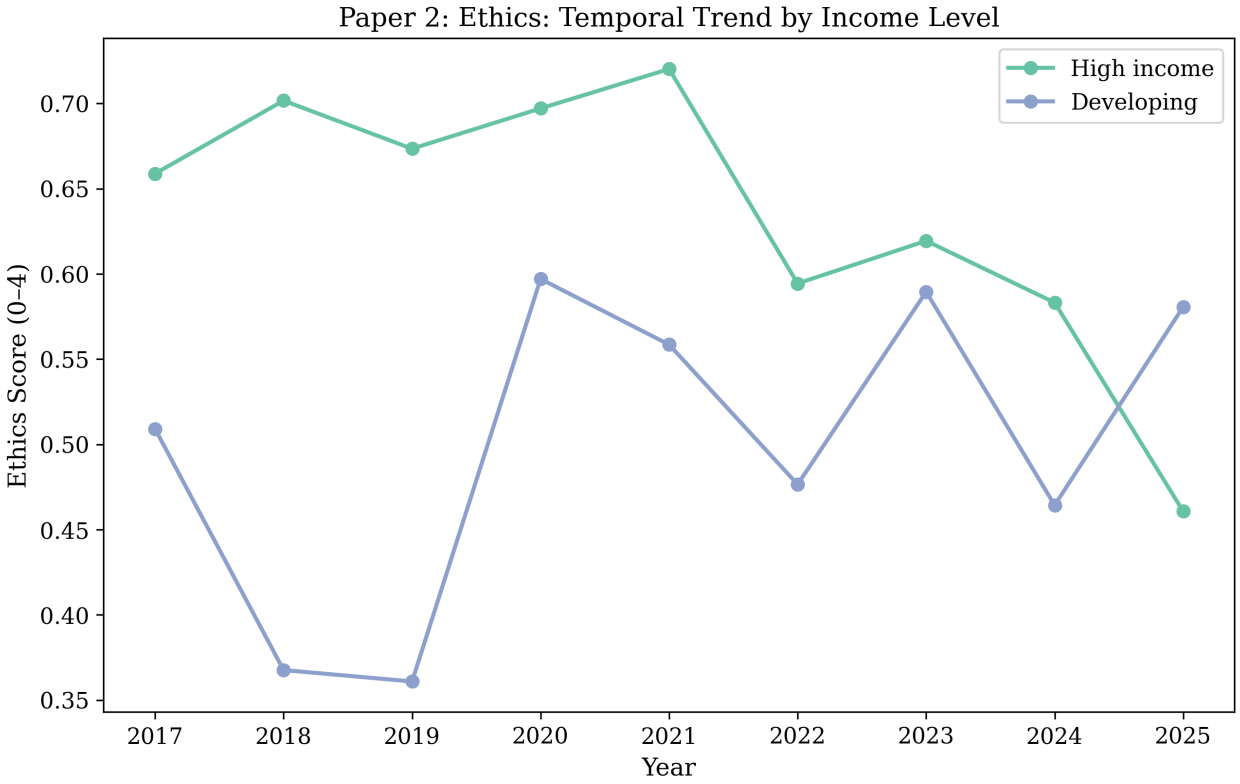

Figure 8.4: Ethics trends by income group.

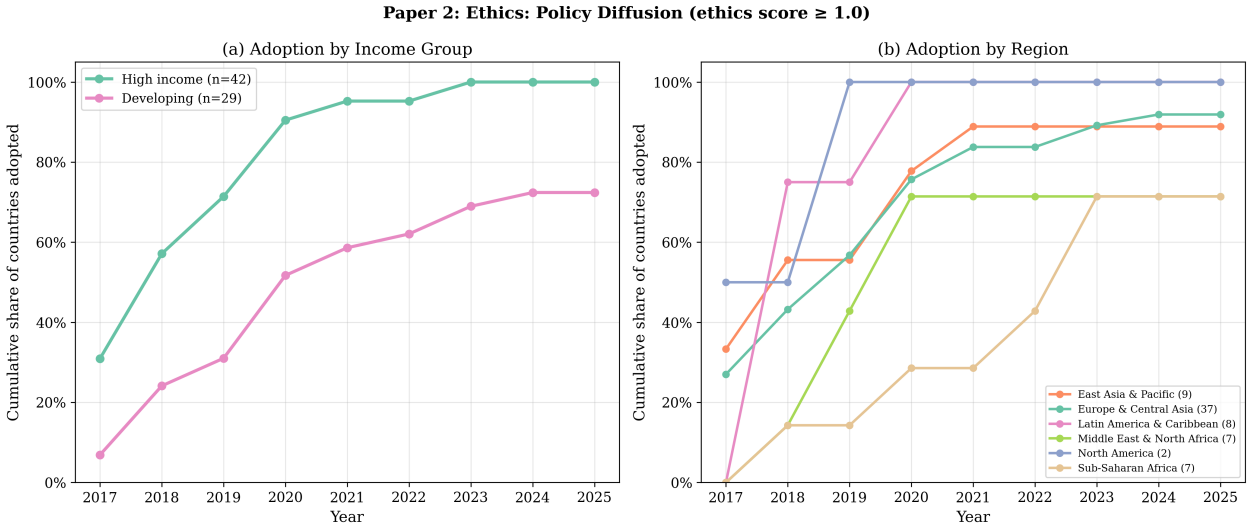

Figure 8.5: Cumulative adoption curves for ethics governance.

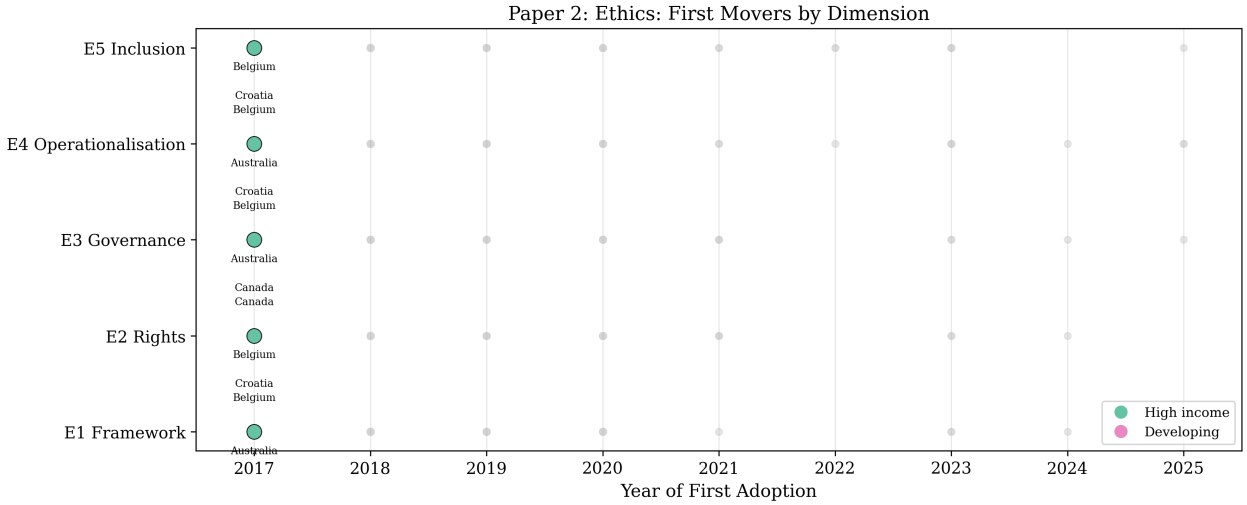

Figure 8.6: First movers in AI ethics governance.

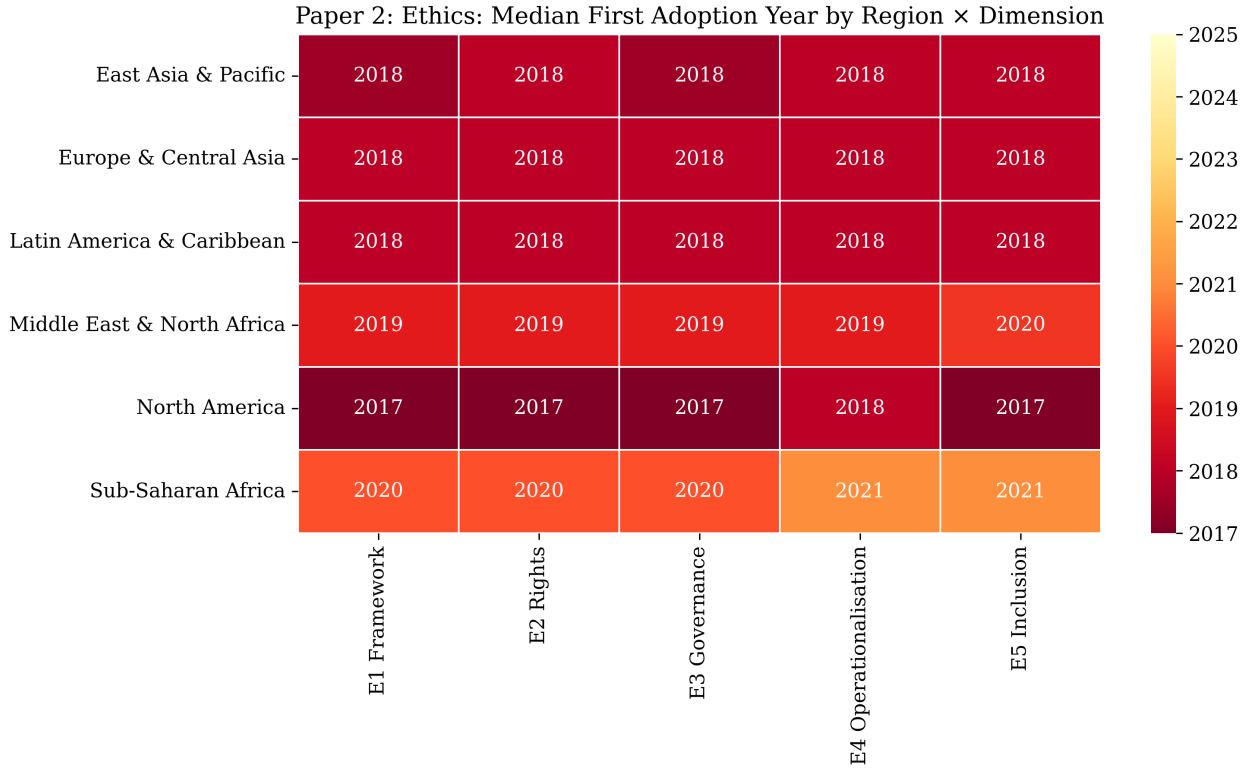

Figure 8.7: Regional diffusion heatmap for ethics policies.

Table 8.2: Ethics diffusion patterns

Metric	Value
HI median first adoption	2018
Developing median first adoption	2020
Adoption lag (HI earlier by)	1.2 years ($p = .021$)
HI adoption by 2025	100%
Developing adoption by 2025	72%
Diffusion direction	98% horizontal

The adoption gap exceeds the quality gap: high-income countries adopted 1.2 years earlier and achieved 100% adoption by 2025 vs. 72% for developing countries. Quality differences are smaller for ethics (d=0.20) than capacity (d=0.30).

Twenty-eight percent of developing countries have no ethics content—complete governance voids without principles, rights protections, or accountability mechanisms. This coverage gap matters more than quality gaps.

Despite this, diffusion is 98% horizontal—developing countries learn from regional peers rather than wealthy countries. The mechanisms of horizontal ethics diffusion differ from capacity diffusion in important ways. Whereas capacity diffusion operates through institutional templates (regulatory structures, agency designs, budget frameworks), ethics diffusion operates through **normative channels**: international human rights instruments, regional ethics declarations, civil society advocacy networks, and academic communities.

International frameworks serve as shared reference points rather than top-down mandates. The UNESCO AI Ethics Recommendation (2021), adopted by 193 member states, provides a common vocabulary and set of principles that countries then adapt to national contexts. Countries do not copy UNESCO's text wholesale; they use it as a legitimating framework to justify domestic ethics governance that reflects local political traditions and rights cultures. The OECD AI Principles play a similar role for member states.

Regional human rights bodies facilitate ethics diffusion within geographic and cultural blocs. The African Commission on Human and Peoples' Rights, the Inter-American Commission on Human Rights, and the ASEAN Intergovernmental Commission on Human Rights all provide institutional channels through which AI ethics norms circulate among peer countries. These bodies embed AI ethics within existing rights protection mandates, making ethics governance an extension of established obligations rather than a novel regulatory burden.

Civil society networks operate across borders within income tiers. Organisations like Access Now, the Web Foundation, and regional digital rights coalitions connect advocates in developing countries who share strategies for embedding rights protections in AI governance. These networks are particularly effective at diffusing E2 (Rights Protection) and E5 (Inclusion) provisions, precisely the dimensions where formal government-to-government channels are weakest.

Academic and diaspora networks connect policy researchers across similar-income countries, enabling knowledge transfer about ethics operationalisation that formal diplomatic channels rarely

address. Researchers who have contributed to ethics frameworks in one country frequently advise neighbouring jurisdictions.

Efficiency Frontier. Capacity analysis showed GDP explains only 3.5% of variation. For ethics, it's even less.

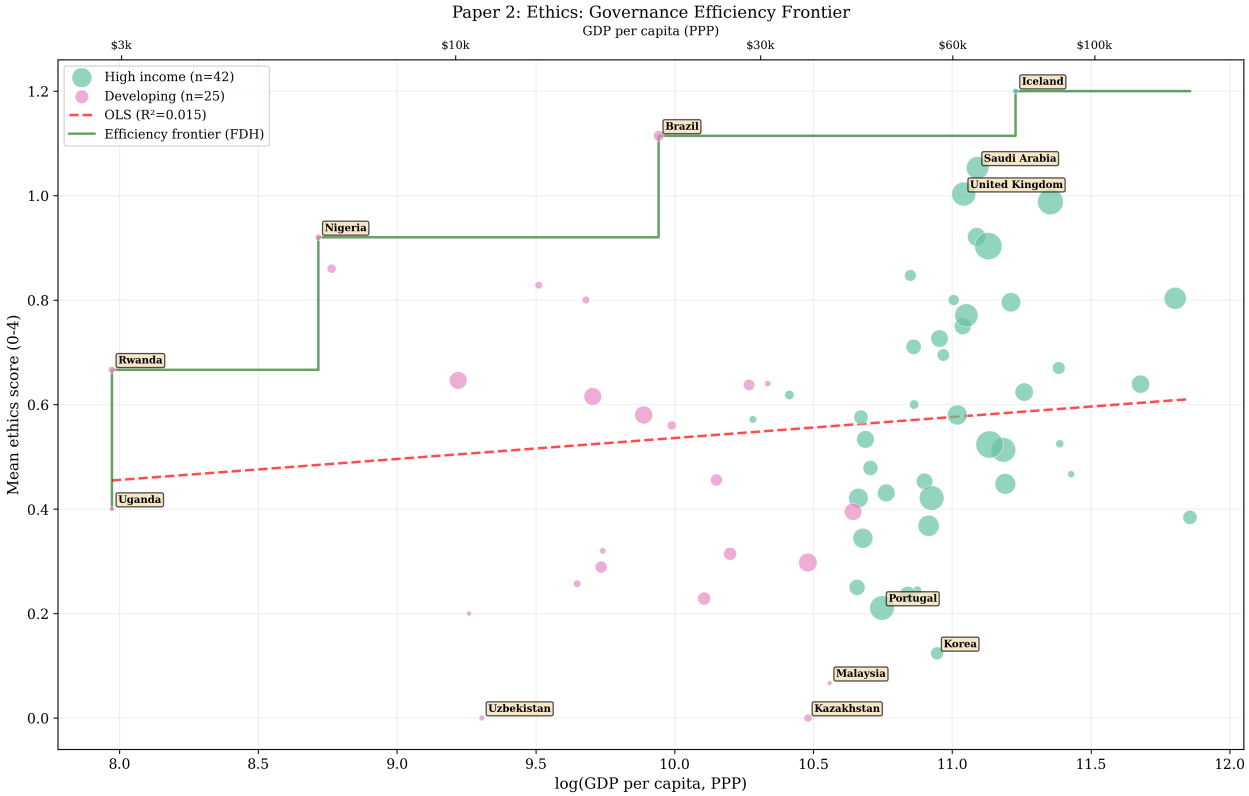

Figure 8.8: Ethics efficiency frontier. GDP explains only 1.5% of country-level ethics variation.

Table 8.3: Ethics efficiency frontier results

Metric	Value
OLS R^2 (score ~ GDP)	0.015
Top overperformer	Iceland (+0.61)
Top underperformer	Kazakhstan (−0.56)
Frontier countries (FDH)	Uganda → Rwanda → Nigeria → Brazil → Iceland
Most efficient (score/$10k GDP)	Rwanda (2.30), Nigeria (1.51)

GDP explains just 1.5% of ethics variation ($R^2 = 0.015$). Knowing GDP tells you almost nothing about ethics quality.

The frontier features developing countries alongside wealthy ones: Uganda, Rwanda, Nigeria, and Brazil all achieve near-maximum ethics at their GDP levels. Rwanda's efficiency ratio of 2.30 ethics

Figure 8.9: Residual ranking for ethics: distance from GDP-predicted score.

points per $10,000 GDP is four times the typical high-income ratio. These countries demonstrate that ethics emerges from political commitment rather than fiscal abundance.

Iceland (+0.61 residual, top overall overperformer) leverages its tradition of deliberative democracy and small-scale governance. Iceland's AI governance framework emerged from extensive public consultation processes—enabled by a population small enough to achieve genuine participatory engagement—and embeds ethics within a constitutional rights tradition that prioritises individual protections. Its data protection authority, though small, exercises meaningful oversight. **Nigeria**'s

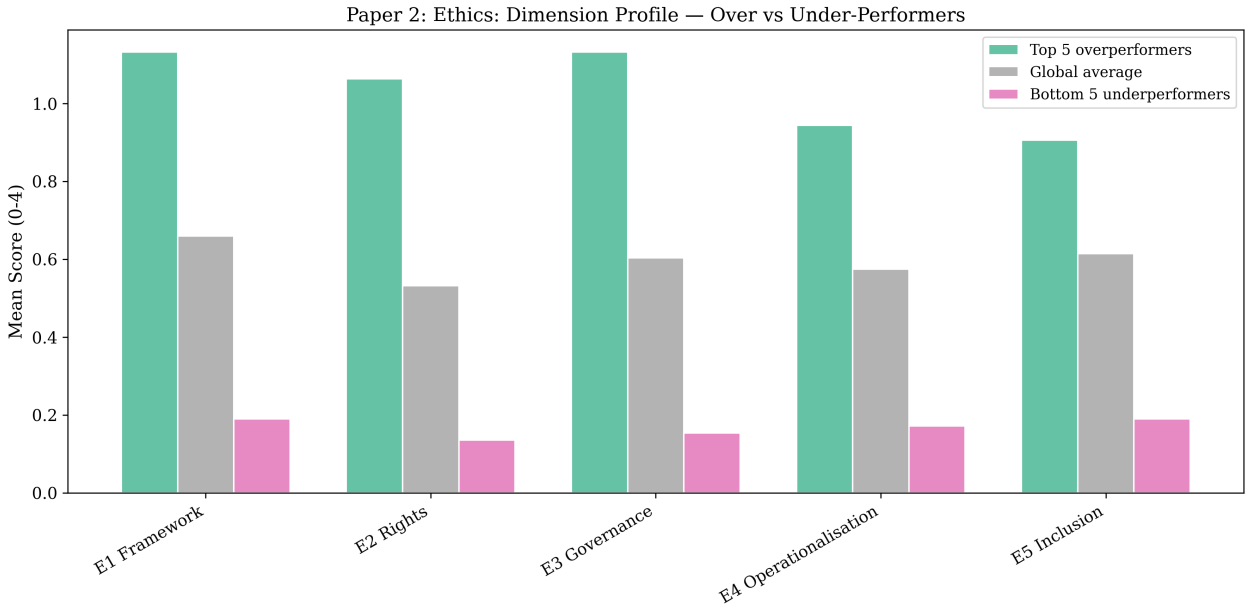

Figure 8.10: Over- and under-performer profiles for ethics.

federal structure creates multiple entry points for civil society engagement with AI governance. The National Information Technology Development Agency (NITDA) developed binding data protection and AI governance frameworks through multi-stakeholder processes that included civil society organisations, academic researchers, and private sector representatives, producing policies that score high on E3 (Governance Mechanisms) and E5 (Inclusion). **Brazil**'s ethics governance builds on decades of participatory governance infrastructure—from participatory budgeting to national policy conferences—that predates AI. The Marco Legal da IA and the broader AI Ethics Framework embed rights protections within this participatory tradition, producing unusually high scores on E2 (Rights Protection) and E5 (Inclusion). Brazil's Conselho Nacional de Proteção de Dados provides institutional continuity for ethics governance.

Kazakhstan underperforms most dramatically (−0.56 residual), with resource wealth failing to generate ethics governance. Several wealthy Asian countries similarly emphasise competitiveness over ethics, producing policies rich in capacity but light on normative commitments.

Diverse pathways work: deliberative democracy (Iceland), federal civil society engagement (Nigeria), participatory governance traditions (Brazil), rights-based constitutionalism (Rwanda). No single institutional model monopolises ethical AI governance.

Summary. Three patterns distinguish ethics dynamics from capacity. First, convergence occurs via decline: high-income countries decline at −0.023/yr ($p = .001$) while developing countries improve modestly. The gap narrows because wealthy countries shift from aspirational frameworks to technical regulation. Second, adoption gaps exceed quality gaps: only 72% of developing countries have any ethics policy (vs. 100% HI). Quality differences are small ($d = 0.20$), but 28% of developing countries have complete ethics voids. Third, GDP explains almost nothing ($R^2 = 0.015$). Rwanda achieves 2.30 ethics points per $10,000 GDP. Political commitment matters; wealth does not.

These findings suggest ethics governance is achievable at any income level. Development interventions should target coverage gaps and facilitate South-South learning from frontier countries like Rwanda, Kenya, and Brazil.

9 Robustness Checks

9.1 How Robust Are Ethics Findings?

Section summary. This section tests ethics findings through text quality restrictions, bootstrap confidence intervals, cluster stability, and sensitivity analyses. A key finding: the income-group ethics gap reverses for well-documented policies.

The preceding sections documented low ethics scores, negligible GDP effects, within-group inequality dominance, and convergence driven by high-income country decline. Before drawing policy conclusions from these findings, this section subjects them to systematic robustness checks. The most consequential test examines the text quality confound: if developing countries' policies appear weaker simply because they are less well documented, then the apparent ethics gap may be a measurement artifact rather than a governance reality.

9.1.1 The Text Quality Confound

All ethics findings depend on a critical assumption: that policy text accurately reflects normative commitments. If text availability varies by income group—with wealthy countries publishing detailed frameworks while developing countries' policies appear as summaries—then apparent ethics gaps may reflect documentation quality rather than ethical governance depth.

The Ethics Gap Reverses.

Table 9.1: Income-group ethics effect by text quality

Sample Restriction	N	Ethics d	Interpretation
All texts	2,097	+0.20***	Small gap
Good-text (500 words)	948	−0.09 (n.s.)	Gap reverses
Excluding stubs	1,754	+0.11 (p=.08)	Marginal

Table 9.1 reveals a notable pattern: the ethics gap (d=0.20) not only disappears but reverses sign (d=-0.09) when analysis is restricted to well-documented policies. Developing countries with adequate documentation slightly outperform wealthy countries on ethics governance, though this reverse gap lacks statistical significance.

The apparent ethics advantage for high-income countries represents a measurement artifact. Developing countries facing documentation challenges nonetheless maintain ethics commitments matching

or exceeding wealthy nations. The full-sample gap (d=0.20) reflects text availability rather than normative sophistication.

Unlike capacity (where the gap merely shrinks to near-zero), ethics shows sign inversion—suggesting developing countries may actually prioritize ethics governance more strongly once measurement quality improves. This aligns with the finding that GDP shows no effect on ethics across all quantiles.

The ethics dimensions (E1 Framework Depth, E2 Rights Protection, E3 Participatory Governance, E4 Operationalization, E5 Inclusion) require normative clarity and political commitment rather than fiscal resources. Developing countries can embed rights protections, establish participatory mechanisms, and articulate ethical principles as effectively as wealthy countries—but shorter available texts obscure this.

9.1.2 Additional Robustness Checks

Bootstrap confidence intervals (1,000 resamples): Ethics d = 0.20 [0.09, 0.30]. The interval is precise but does not address validity. For the good-text subsample, the 95% CI spans [−0.22, +0.04], comfortably including zero and confirming that the apparent income-group gap lacks statistical significance once documentation quality is controlled.

Cluster stability: The two-cluster solution ("Ethics-Light" vs "Ethics-Engaged") is optimal. Silhouette scores decline monotonically: 0.42 for k=2, dropping for higher k. The binary typology reflects a genuine bimodal structure—policies either engage with ethics comprehensively or ignore it almost entirely. This bimodality is consistent with the severe floor effect (36.3% zeros) and suggests that ethics governance is closer to a binary commitment decision than a continuous gradient.

Sensitivity tests (full details in the *Robustness Appendix* of the companion Methods volume):

- *Excluding international organisations*: Removing the 119 policies from the EU, OECD, UN, and multilateral bodies leaves all results unchanged. The ethics findings reflect national governance patterns rather than the influence of international frameworks, which tend to score high on principles but may operate under different institutional logics.
- *Ordinal regression*: Treating ethics scores as ordered categories rather than continuous variables preserves the rank ordering of all predictors. The zero GDP effect, within-group inequality dominance, and convergence pattern all hold, confirming that findings do not depend on the interval-scale assumption.
- *Winsorizing extremes*: Trimming the top and bottom 5% of scores yields coefficients within 10% of baseline estimates, ruling out the possibility that outlier policies—either extremely sophisticated frameworks or unusually empty ones—drive the results.
- *Alternative income classifications*: The zero GDP effect persists across two-group (HI vs developing), three-group (HI/UMI/LMI+LI), and four-group (full World Bank tiers) specifications. The non-relationship between wealth and ethics governance is not an artefact of how the income divide is drawn.
- *Text quality thresholds (300–1,000 words)*: The income-group gap disappears at all thresholds 500 words and reverses sign at 1,000 words. This dose-response pattern—stronger reversal

with more stringent quality thresholds—provides compelling evidence for the measurement-artefact interpretation.
- *Temporal subsamples*: The convergence pattern (gaps narrowing via high-income decline) holds in both the 2017–2020 and 2021–2025 subsamples, confirming that the trend is not driven by a single policy wave or by the composition of newer policies alone.

9.1.3 Ethics vs Capacity Asymmetry and Summary

The text quality confound affects ethics differently than capacity:

Table 9.2: Text quality effects on capacity vs ethics

	Capacity	Ethics
Full sample	d = +0.30***	d = +0.20***
Good text	d = +0.04 (n.s.)	d = −0.09 (n.s.)
Change	−87% shrinkage	Sign reversal

Capacity shows shrinkage (the gap remains positive but becomes trivial). Ethics shows reversal (developing countries outperform once measured properly). This asymmetry is interpretable: capacity requires infrastructure that wealth facilitates, while ethics requires commitments independent of GDP.

Summary.

Table 9.3: Ethics robustness summary

Finding	Robust?	Caveat
Income-group ethics gap	Artifact	Reverses for good texts
GDP zero ethics effect	Yes	Robust across all quantiles
Within-group inequality (99%)	Yes	All specifications
Convergence (gaps narrowing)	Yes	Temporal pattern robust
Horizontal diffusion	Yes	Consistent pattern

The income-group ethics gap appears to be a measurement artifact—developing countries' ethics commitments match or exceed wealthy countries when documentation quality is equivalent. This represents one of the study's more optimistic findings about global ethics governance.

10 Discussion

10.1 Implications for Ethics Governance

The empirical analysis has established four central findings: ethics governance is weak globally, GDP does not predict ethics depth, nearly all variation occurs within rather than between income groups, and the ethics gap is narrowing over time. The robustness checks confirmed that the apparent income-group gap largely reflects documentation quality rather than genuine governance differences. This section draws out the implications of these findings for theory, policy, and research practice.

10.1.1 Ethics Without Wealth

GDP shows no effect on ethics scores at any quantile—a finding that holds across OLS, quantile, multilevel, and Tobit specifications. This contradicts the implicit assumption in much of the AI governance literature that ethical sophistication follows economic development. Lee (2018)'s framing of developing countries facing ethical "catch-up" challenges implies a sequence—build economic capacity, then layer ethics on top—that the data do not support. Kenya's AI and Data Protection regulations, Brazil's AI Ethics Framework, Colombia's AI strategy, and Rwanda's governance instruments all demonstrate high ethics scores despite modest GDP levels.

The mechanism is interpretable: unlike implementation capacity, which requires fiscal resources, ethical governance requires political will, democratic traditions, and civil society strength—factors that operate independently of GDP. Gwagwa et al. (2020)'s documentation of indigenous African ethics frameworks and Müller (2021)'s analysis of the UNESCO AI Ethics Recommendation suggest that normative pathways are plural rather than sequential. Approximately 97% of policies score below 2.0 on operationalisation, a deficit affecting high-income and developing countries alike.

The documentation confound reinforces this conclusion. The income gap ($d = 0.20$) reverses direction for well-documented policies ($d = -0.09$, n.s.)—when documentation quality is comparable, developing countries slightly outperform wealthy countries on ethics governance. Any study using policy text as data risks confounding documentation practices with substantive governance. Researchers should routinely stratify by text quality and report results separately.

10.1.2 The Regulation-Replacing-Aspiration Paradox

Ethics gaps are narrowing between income groups—but through an unexpected mechanism. High-income countries are *declining* (-0.023/yr, $p = .001$) while developing countries remain roughly stable ($+0.016$/yr, n.s.). Is this convergence good news or bad news?

The optimistic interpretation: convergence reflects governance maturation. Early high-income frameworks (Canada's Pan-Canadian AI Strategy, France's Villani Report, Germany's AI Strategy) were deliberative documents that naturally scored high on ethical articulation because that was their primary purpose. Later instruments (the EU AI Act, sector-specific regulations, procurement standards) assume those ethical foundations and focus on compliance infrastructure. The decline in ethics scores reflects a shift from *articulating* principles to *implementing* them—precisely the transition that Hagendorff (2020) and Selbst et al. (2019) called for.

The pessimistic interpretation: convergence through decline means the global stock of ethical guidance is thinning. If later policies assume ethical foundations without restating or updating them, those foundations may erode as institutional memory fades, staff turn over, and political priorities shift. The founding frameworks become legacy documents that subsequent implementers may never read. What Whittaker et al. (2018) called the gap between commitment and practice may widen even as the gap between countries narrows.

The truth likely lies between these interpretations. The pattern calls for layered governance architectures: aspirational frameworks that establish and periodically refresh ethical foundations, complemented by binding regulations that operationalise principles into enforceable requirements. Countries that allow aspirational frameworks to atrophy while producing only technical regulations risk losing the normative anchors that give technical governance its purpose.

10.1.3 Operationalisation and International Organisations

E4 Operationalisation scores lowest among ethics dimensions. Policies invoke principles without specifying compliance requirements, enforcement procedures, or technical standards. What Selbst et al. (2019) termed "fairness gerrymandering"—proclaiming commitments without operational definitions—characterises much of global ethics governance. Fjeld et al. (2020)'s binary coding approach would miss this pattern entirely: a policy mentioning "transparency" and one establishing mandatory algorithmic impact assessments would both score as present.

Addressing this gap requires four concrete steps: specifying operational definitions (what does fairness mean in public procurement versus criminal justice?), establishing compliance procedures (who certifies compliance and how?), creating enforcement mechanisms (what sanctions apply for non-compliance?), and providing technical guidance (what audit methodologies satisfy transparency requirements?). International organisations can play a catalytic role by developing operational toolkits—impact assessment templates, audit standards, compliance checklists—that translate principles into implementable requirements.

Implications for International Organisations. These findings speak directly to the work of UNESCO, the OECD, and regional bodies. The UNESCO AI Ethics Recommendation (2021), adopted by 193 member states, provides a shared normative framework—but this study shows that framework articulation is precisely where countries already perform best. What countries need is operationalisation support: practical guidance on translating UNESCO's principles into national legislation, compliance mechanisms, and enforcement procedures. The OECD's ongoing governance monitoring likewise captures policy adoption but not operationalisation depth; supplementing adoption tracking with the kind of dimensional scoring used here would provide more actionable intelligence.

Regional bodies (the African Union, ASEAN, the Red Iberoamericana de Inteligencia Artificial) are well positioned to facilitate the horizontal diffusion that the data show dominates ethics governance. Rather than top-down technical assistance from wealthy-country frameworks, South-South peer learning—connecting Kenya with Nigeria, Brazil with Colombia, Rwanda with Uganda—appears more consequential for ethics governance development.

10.1.4 Limitations

Five limitations qualify these findings. **First**, the study measures ethics governance as expressed in policy text, not as experienced by affected populations. A policy scoring 4/4 on rights protection may fail to protect rights in practice; a policy scoring 0 may operate within a legal tradition that provides rights protections through other instruments. **Second**, LLM-based scoring may carry systematic biases. All three models were trained on corpora overrepresenting English-language, Western governance documents; they may interpret non-Western governance traditions through Western normative frameworks, potentially underscoring indigenous approaches that achieve ethical governance through unfamiliar institutional forms. **Third**, the OECD.AI Observatory's coverage skews toward OECD member states and their partners, likely understating governance activity in countries with limited international reporting capacity. **Fourth**, the cross-sectional design cannot establish causal relationships—GDP's non-effect on ethics could reflect reverse causation (countries with strong ethics traditions invest differently in economic development) or omitted variables (democratic traditions driving both). **Fifth**, the 77% high-income composition of the corpus may underpower developing-country analyses; the non-significant reversal ($d = -0.09$) for well-documented policies might reach significance with a larger developing-country sample.

11 Conclusion

11.1 Toward Operationalised Ethics Governance

How deeply do AI policies embed ethical commitments? The evidence indicates: not very deeply. The modal policy scores below 2/4 on operationalisation, invoking principles without governance depth. Over 36% of all policies contain zero ethics content, and 97% fall below the scale midpoint.

Yet the distributional patterns offer grounds for cautious optimism.

11.1.1 Main Findings

Ethics governance is not contingent on economic development. GDP shows no effect on ethics scores at any quantile. Countries at any income level can embed rights protections, establish participatory mechanisms, and operationalise ethical principles. Kenya, Brazil, Colombia, Rwanda, Iceland, and Nigeria all demonstrate this, achieving ethics scores that exceed GDP predictions. This contradicts the "luxury good" narrative that treats ethical governance as a post-development achievement.

Gaps are narrowing, through an unexpected mechanism. Developing countries have not caught up to high-income standards; rather, high-income countries have declined as governance shifts from aspirational frameworks to technical regulation. If trends continue, the gap closes entirely by 2030. Whether this convergence is encouraging (maturation) or concerning (erosion of normative foundations) depends on whether later technical policies successfully operationalise the principles that earlier frameworks articulated.

Within-group inequality dominates. Approximately 99.5% of ethics variation occurs within income groups—even more extreme than capacity's 98.8%. The conventional framing of a global North-South ethics divide obscures the reality that political will, democratic traditions, and civil society strength—not national wealth—drive ethics governance depth.

Operationalisation remains the binding constraint. Convergence on principle *mention* means little without compliance mechanisms and enforcement capacity. E4 Operationalisation scores lowest among dimensions, and the gap between articulating principles and implementing them characterises global AI governance regardless of income level.

Documentation quality confounds cross-national comparison. The income-group gap (d = 0.20) reverses sign (d = −0.09) for well-documented policies. Developing countries' true ethics commitments likely exceed what available texts suggest. This finding carries methodological implications for any text-based governance comparison.

11.1.2 Future Research and the Observatory Vision

Several directions warrant investigation. **First**, validating LLM ethics scores against human expert coding would test whether automated scoring captures the nuances of ethical governance that rubric-based assessment intends to measure. **Second**, tracking the same jurisdictions over time would enable within-country panel analysis, testing whether specific policy interventions produce measurable improvements. **Third**, linking policy text scores to implementation outcomes would address the fundamental question this study cannot: do policies that score high on ethics governance actually produce better outcomes for affected populations? **Fourth**, extending analysis to sector-specific policies would test whether ethics governance varies across application domains. **Fifth**, the convergence-through-decline finding motivates research on normative sustainability: do countries that shift from aspirational to technical governance successfully embed ethical principles in implementation?

The findings presented here are descriptive and correlational. The non-effect of GDP on ethics could reflect reverse causation, omitted variables (democratic maturity driving both), or genuine causal independence. Establishing causal relationships would require quasi-experimental designs exploiting exogenous shocks to governance or instrumental variable approaches.

The Observatory Vision. The research infrastructure developed here—the scraping pipeline, text extraction tools, LLM scoring framework, and analytical code—supports a living observatory of AI ethics governance. Practical applications include annual scoring rounds tracking how ethics governance evolves as countries revise policies and adopt new instruments; country-level ethics scorecards providing dimension-specific benchmarks against regional and income-group peers; operationalisation toolkits translating abstract principle scores into concrete improvement recommendations; and open-source benchmarking tools enabling governments, civil society organisations, and international bodies to assess their own policies against the global distribution.

The observatory model is particularly valuable for tracking whether the convergence trend continues—and whether it ultimately reflects governance maturation or normative erosion.

Code, data, and methods: https://github.com/lsempe77/ai-governance-capacity

---

Ethics governance ultimately reflects political commitment to translate values into enforceable requirements—building infrastructure that makes principles meaningful.

European Parliament and Council. 2024. "Regulation (EU) 2024/1689 Laying down Harmonised Rules on Artificial Intelligence (AI Act)."

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI." Berkman Klein Center, Harvard University.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. "AI4Peopleâ an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28: 689–707.

Gwagwa, Arthur, Emre Kazim, Abeba Hilliard, Kathleen Siminyu, and Mercy Smith. 2020. "Bridging the Gap in AI Implementation in African Healthcare: An Ubuntu Framework for Responsible AI." *Information* 11 (9): 423. https://doi.org/10.3390/info11090423.

Hagendorff, Thilo. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30: 99–120.

Hildebrandt, Mireille. 2019. "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning." *Theoretical Inquiries in Law* 20 (1): 83–121. https://doi.org/10.1515/til-2019-0004.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–99.

Lee, Kai-Fu. 2018. "AI Superpowers: China, Silicon Valley, and the New World Order." *Houghton Mifflin Harcourt*.

Mittelstadt, Brent. 2019. "Principles Alone Cannot Guarantee Ethical AI." *Nature Machine Intelligence* 1: 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Müller, Vincent C. 2021. "Ethics of Artificial Intelligence and Robotics." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/ethics-ai/.

OECD. 2019. "OECD Principles on Artificial Intelligence."

Rahwan, Iyad. 2018. "Society-in-the-Loop: Programming the Algorithmic Social Contract." *Ethics and Information Technology* 20: 5–14. https://doi.org/10.1007/s10676-017-9430-8.

Rességuier, Anaïs, and Rowena Rodrigues. 2020. "AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics." *Big Data & Society* 7 (2). https://doi.org/10.1177/2053951720942541.

Selbst, Andrew D., danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. "Fairness and Abstraction in Sociotechnical Systems." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598.

Stilgoe, Jack. 2020. "Who's Driving Innovation? New Technologies and the Collaborative State." *Palgrave Macmillan*. https://doi.org/10.1007/978-3-030-32320-2.

UNESCO. 2021. "Recommendation on the Ethics of Artificial Intelligence."

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. "AI Now Report 2018." AI Now Institute, New York University. https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12 (4): 505–23. https://doi.org/10.1111/rego.12158.