UNESCO AI Ethics Recommendation

Global Alignment Assessment Across 2,100+ Policies

Lucas Sempé

February 11, 2026

Table of contents

1 UNESCO AI Ethics Recommendation: Global Alignment Assessment

Measuring Policy Convergence with UNESCO Principles Across 2,100+ Documents

Preface

In November 2021, UNESCO achieved a multilateral milestone: 193 countries adopted a shared framework for AI ethics. This study examines subsequent policy implementation.

This study analyzed 2,100+ AI policies from 70+ jurisdictions to measure the extent to which national governance aligns with UNESCO's ten core values and eleven action areas. The analysis reveals partial adoption. Countries selectively emphasize components that match existing priorities while largely neglecting environmental sustainability, gender, and culture.

1.1 What We Found

Mean alignment sits at 1.68 out of 4.0—somewhere between "mentioned" and "described," well short of operationalized. Human rights and transparency score highest; sustainability trails far behind. Only 28% of policies show comprehensive alignment; the remaining 72% adopt selectively. Post-2021 policies show modest improvement, but the 2021 milestone changed rhetoric more than practice.

1.2 How We Measured It

The study employed three large language models (Claude Sonnet 4, GPT-4o, Gemini Flash 2.0) as automated policy analysts, scoring each document against UNESCO's 21 components. The ensemble achieves excellent reliability (ICC = 0.827)—comparable to human expert agreement on complex policy assessment tasks.

Citation: Sempé, L. (2026). *UNESCO AI Ethics Recommendation: Global Alignment Assessment.* International Initiative for Impact Evaluation (3ie).

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

2 Introduction

2.1 Global Alignment with UNESCO AI Ethics

In November 2021, 193 countries adopted a shared vision for ethical AI. This study examines whether countries translated this international commitment into national policy.

2.1.1 The UNESCO Milestone

The **Recommendation on the Ethics of Artificial Intelligence**, adopted in November 2021, stands as the first global normative instrument on AI ethics. It is ambitious in scope: ten core values (from human rights to environmental sustainability) and eleven action areas (spanning ethical impact assessment to gender policy). No previous multilateral agreement had attempted anything this comprehensive for AI governance. This study examines the extent to which countries have translated this ambition into national policies.

2.1.2 Research Questions

This study addresses three research questions: How closely do national AI policies align with UNESCO's framework? Which values receive priority and which are neglected? Has the Recommendation influenced policy development since 2021?

2.1.3 Measurement Framework

We score each policy on 21 UNESCO components (10 values + 11 action areas) using the same LLM ensemble methodology achieving ICC = 0.827. Scores reflect depth of engagement: 0 = absent, 1 = mentioned, 2 = described, 3 = operationalized, 4 = comprehensive.

The short answer: partial adoption. Mean alignment sits at 1.68 out of 4.0—somewhere between "mentioned" and "described" but well short of "operationalized." Countries cherry-pick. Human rights and transparency score highest (1.92 and 1.85); environmental sustainability trails at 1.28. Only 28% of policies show comprehensive alignment; the remaining 72% engage selectively. Post-2021 policies do show stronger alignment, though the improvement is modest rather than transformative.

The remainder of this book proceeds as follows. Section 5.1 maps the alignment landscape—which UNESCO components countries address and how deeply. Section 6.1 investigates what drives alignment, revealing that national wealth is not a significant determinant. Section 7.1 identifies four

distinct policy archetypes through cluster analysis. Section 8.1 examines whether the 2021 adoption date actually changed anything. Section 9.1 stress-tests our findings, and Section 10.1 and Section 11.1 draw out the implications.

3 Literature Review

3.1 Theoretical Foundations

When 193 countries adopt a recommendation, several policy diffusion mechanisms may operate. The international relations literature suggests several possibilities—coercion, competition, learning, legitimacy-seeking—each implying different patterns of adoption. Understanding these mechanisms helps explain why some UNESCO components spread faster than others.

However, the norm diffusion literature confronts an identification problem: observing policy convergence does not reveal which mechanism drove it. Countries might adopt UNESCO language because they genuinely learned from multilateral deliberation (learning), because they seek international legitimacy without implementation commitment (emulation), or because major powers condition benefits on adoption (coercion). Distinguishing these mechanisms requires examining not only *whether* countries adopt UNESCO components but *how deeply* they implement them and *which* components receive priority—patterns this study traces through systematic measurement.

3.1.1 The UNESCO Recommendation

The **Recommendation on the Ethics of Artificial Intelligence**, adopted in November 2021, established ten core values (spanning human rights to environmental sustainability) and eleven action areas (from ethical impact assessment to regulation). Taddeo and Floridi (2021) characterizes it as "soft law"—non-binding but influential through moral authority and coordination effects. Floridi et al. (2021) argues it provides the most comprehensive multilateral AI ethics framework to date, integrating principles from diverse philosophical traditions.

Yet comprehensiveness may come at a cost: a framework spanning 10 values and 11 action areas risks becoming unwieldy, allowing countries to claim alignment through selective adoption while neglecting components that conflict with domestic priorities. The breadth that makes UNESCO inclusive may simultaneously weaken its normative force. Moreover, soft law's influence depends on domestic actors who care about international reputation and coordination—mechanisms that may operate differently in democracies versus autocracies, or in countries deeply embedded in international institutions versus those with limited multilateral engagement. These heterogeneous responses become empirically testable through systematic alignment measurement.

3.1.2 International Norm Diffusion

Simmons, Dobbin, and Garrett (2006) identifies four mechanisms driving policy convergence: coercion (powerful actors impose rules), competition (regulatory arbitrage pressures jurisdictions toward common standards), learning (countries emulate successful policies from peers), and emulation (mimetic isomorphism driven by legitimacy rather than function). This study examines which mechanisms drive UNESCO alignment. If coercion or competition dominates, wealth-driven adoption patterns should emerge. If learning or emulation dominates, horizontal diffusion within regions or income groups should occur. However, Simmons and colleagues' framework was developed for economic policy domains (trade liberalization, capital account opening) where material incentives and competitive pressures are pronounced. Whether the same mechanisms operate in AI ethics—a domain characterized by normative rather than material stakes, weak enforcement mechanisms, and limited regulatory arbitrage opportunities—remains uncertain. The absence of wealth-driven adoption patterns documented in Section 18 challenges coercion and competition mechanisms, suggesting learning or legitimacy-seeking may dominate UNESCO diffusion.

The Brussels Effect provides a complementary lens. Bradford (2020) theorizes that EU regulations become global standards through market power; in data protection, GDPR created de facto global norms as firms adopted EU standards globally rather than maintaining separate compliance systems. But AI ethics differs from data protection: no single jurisdiction dominates global AI markets sufficiently to impose standards unilaterally. UNESCO's multilateral framework thus operates through different diffusion mechanisms than GDPR, emphasizing learning and legitimacy over market coercion. Moreover, Bradford's analysis focuses on regulatory domains where firms face compliance costs that incentivize convergence. AI ethics principles impose fewer direct compliance costs—many guidelines remain aspirational without enforcement mechanisms. This reduces firms' incentives to lobby for global harmonization, potentially weakening market-driven diffusion. The selective adoption patterns documented in Section 17 support this interpretation: countries adopt UNESCO's normative vocabulary without implementing its governance requirements, suggesting legitimation rather than market-driven convergence.

3.1.3 Global Standards and Local Adaptation

Acharya (2004) distinguishes **localization** (modifying global norms to fit local contexts) from **transplantation** (wholesale adoption). Wiener and Puetter (2020) shows that international norms rarely transplant unchanged; they undergo reinterpretation reflecting local values, institutions, and priorities.

The measurement framework captures this by scoring **depth of engagement** rather than binary adoption: policies can mention UNESCO values (score 1), describe them contextually (score 2), operationalize them through requirements (score 3), or establish comprehensive governance (score 4). This approach distinguishes superficial from substantive adoption.

However, the localization literature faces a conceptual tension: if norms must adapt to local contexts to gain traction, how do we identify "the norm" being diffused? When a country mentions "human dignity" but interprets it through religious rather than rights-based frameworks, has UNESCO's human dignity value been adopted, rejected, or localized? The cluster analysis in Section 19 reveals

that countries engage with UNESCO through distinct archetypes—comprehensive aligners, moderate aligners, selective aligners, and minimal engagement—suggesting that localization operates not through uniform adaptation but through divergent pathways of selective emphasis.

3.1.4 Value Priorities and Selectivity

Jobin, Ienca, and Vayena (2019) documents convergence on core AI ethics principles but divergence on prioritization. Winfield and Jirotka (2021) shows regional variation: European frameworks emphasize human rights and fundamental freedoms, Asian frameworks emphasize social harmony and collective welfare, Middle Eastern frameworks emphasize cultural values and religious principles.

UNESCO's framework accommodates this diversity through breadth, including 10 values and 11 action areas, which allows countries to prioritize different components while claiming UNESCO alignment. This study examines whether countries adopt the full framework or selectively emphasize values matching existing priorities.

Yet the convergence-with-divergence pattern creates measurement challenges: if countries adopt common principle *labels* (fairness, transparency, accountability) while diverging on operational *meanings*, does this constitute normative convergence or merely semantic convergence? The item-level analysis in Section 17 reveals that coverage rates vary dramatically across UNESCO components—some exceed 80% adoption while others fall below 10%—suggesting that countries converge on politically safe principles while avoiding components that would require substantial governance reform or challenge dominant interests.

3.1.5 Implementation Challenges

Cihon, Maas, and Kemp (2021) documents the obstacles to translating UNESCO principles into national governance: the values are vague and require national specification; 21 components create a substantial implementation burden; action areas like "ethical impact assessment" lack established templates; and multiple government agencies must coordinate on different areas.

These challenges predict variable UNESCO alignment even among committed member states. The measurement approach distinguishes rhetoric (mentioning UNESCO) from implementation (operationalizing UNESCO components).

Cihon and colleagues' analysis, however, examines implementation challenges conceptually without measuring their severity empirically. Which implementation obstacles matter most? Do coordination challenges (requiring inter-agency agreement) or specification challenges (requiring technical detail) pose greater barriers? The finding in Section 20 that post-2021 policies show only modest alignment increases despite UNESCO providing a detailed framework suggests specification challenges may be less constraining than political economy factors—countries may understand *how* to implement UNESCO but lack *incentives* to do so without enforcement mechanisms or domestic constituencies demanding compliance.

3.1.6 Research Questions and Contribution

This literature motivates three research questions: How aligned are national policies with UNESCO's 21-component framework? Which values and action areas receive priority, and which are neglected? Has alignment increased since 2021?

These questions address a broader theoretical puzzle: under what conditions do international normative instruments influence domestic governance? The UNESCO case provides analytical leverage because its 2021 adoption creates a temporal discontinuity, its comprehensive framework enables component-level analysis, and its global participation permits cross-national comparison. If soft law influences governance, we should observe post-2021 alignment increases, priority given to enforcement-backed components, and adoption concentrated among countries embedded in multilateral institutions. The empirical patterns test these expectations.

Contribution. This study addresses three gaps in the norm diffusion literature. **First**, while scholars have theorized soft law influence mechanisms, systematic measurement of soft law adoption patterns across large samples remains rare. This study scores UNESCO alignment across 2,100+ policies, enabling statistical analysis of adoption determinants and temporal dynamics. **Second**, the localization literature emphasizes how norms adapt to local contexts, yet provides limited evidence on *which* norm components undergo localization versus transplantation. The component-level analysis distinguishes broadly adopted versus selectively emphasized UNESCO items, revealing that countries converge on abstract values while diverging on operational requirements. **Third**, existing studies examine whether international instruments influence domestic policy without testing whether influence operates through changed discourse or changed governance structures. The depth-based scoring approach distinguishes mention from operationalization, showing that UNESCO changed what policies *discuss* substantially more than what they *implement*.

The findings challenge optimistic accounts of soft law influence: the post-2021 alignment increase, while statistically significant, leaves policies below the operationalization threshold; selective adoption patterns suggest legitimation rather than genuine normative internalization; and the absence of wealth effects indicates that UNESCO diffuses through horizontal emulation rather than through capacity-building or coercion. These patterns are consistent with "decoupling"—countries adopt international norms symbolically while maintaining existing governance arrangements.

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 Scoring Framework

Each policy was scored on the same 10-dimension capacity-ethics framework used across all three companion studies (see Book 4 for full rubric and validation): five capacity dimensions (C1–C5) and five ethics dimensions (E1–E5), each on a 0–4 scale. Composite scores are unweighted means.

4.1.2 UNESCO Alignment Scoring

In addition to the shared 10-dimension framework, this study employs a **UNESCO-specific alignment assessment**. Each policy was scored on **25 UNESCO components** drawn from the Recommendation on the Ethics of Artificial Intelligence: 4 values (human rights & dignity, living in peaceful societies, diversity & inclusiveness, environment & ecosystem flourishing), 10 principles (proportionality, safety & security, fairness, transparency, responsibility, privacy, human oversight, sustainability, awareness & literacy, multi-stakeholder governance), and 11 policy action areas (ethical impact assessment, ethical governance, data policy, development & international cooperation,

environment, gender, education & research, health, economy, culture, and communication & information).

For each component, the LLM ensemble assessed two metrics: **coverage** (binary: does the policy mention this component?) and **depth** (1–5 scale: word-level mention, sentence-level engagement, paragraph-level treatment, section-level analysis, or comprehensive integration). The composite UNESCO alignment score (0–100) weights coverage breadth at 60% and normalised depth quality at 40%, capturing both *whether* a policy addresses a component and *how seriously* it engages with it.

> **Note**
>
> **Two distinct metrics.** This book uses two UNESCO-related metrics. The **10-dimension capacity/ethics composite** (0–4 scale) measures general governance quality using the C1–C5 and E1–E5 framework shared across all three studies. The **25-item UNESCO alignment score** (0–100 scale) measures specific coverage of and depth on UNESCO's own framework components. Both are valid; they measure different things.

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

5 UNESCO Alignment Landscape

5.1 The Alignment Landscape: Coverage and Depth

The extent to which AI policies address UNESCO's 25-item framework depends on the evaluation criterion. Breadth coverage appears moderately encouraging—policies mention approximately half the UNESCO items on average. However, depth analysis reveals a substantial gap: proclamation far outpaces implementation.

5.1.1 Overall Alignment Score Distribution

The composite **UNESCO alignment score** (0–100) weights coverage breadth at 60% and normalized depth quality at 40% across all 25 UNESCO items. This metric captures both *whether* a policy mentions a component and *how seriously* it engages with it.

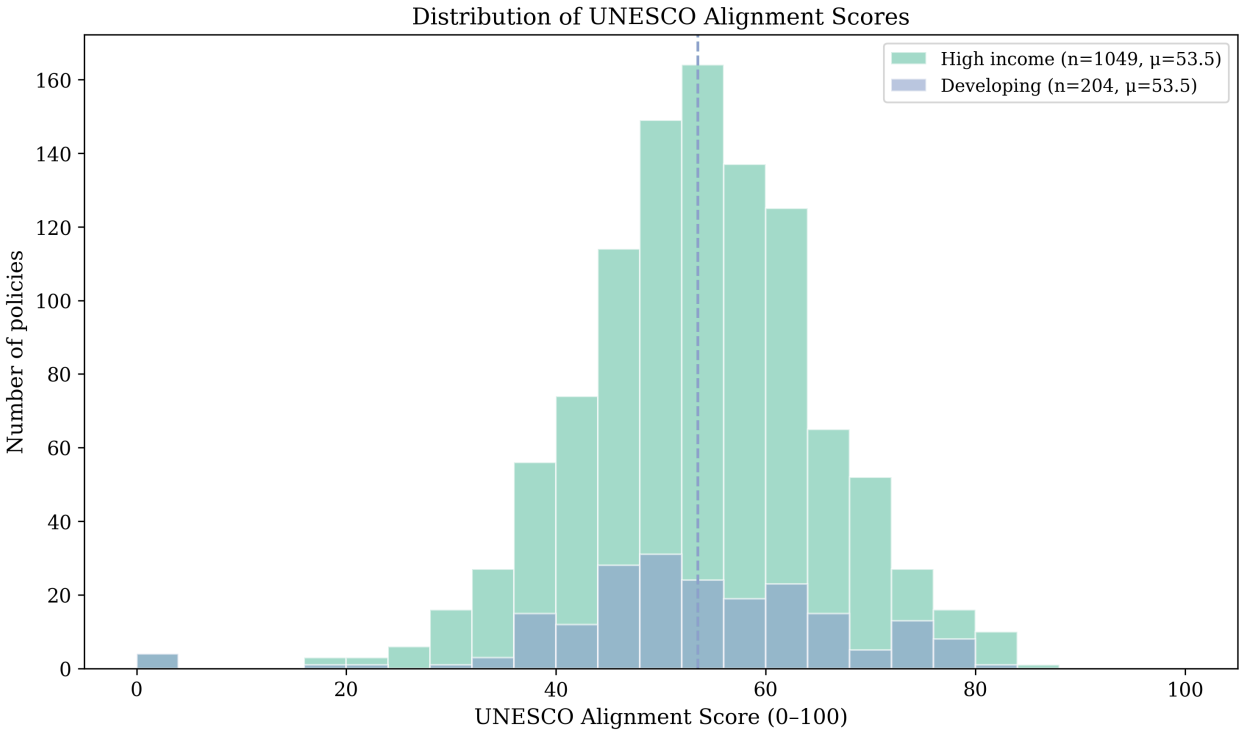

Figure 5.1: Distribution of UNESCO alignment scores across 1,326 AI policies. The distribution is approximately normal, centred on 54 with moderate spread.

The mean is **53.9** (SD = 12.2, median = 53.8). No policy achieves full alignment—the maximum approaches 85, the minimum near 20. Most policies engage with approximately half the framework at moderate depth. However, "moderate" engagement indicates substantial room for improvement.

5.1.2 Coverage Across the 25 UNESCO Items

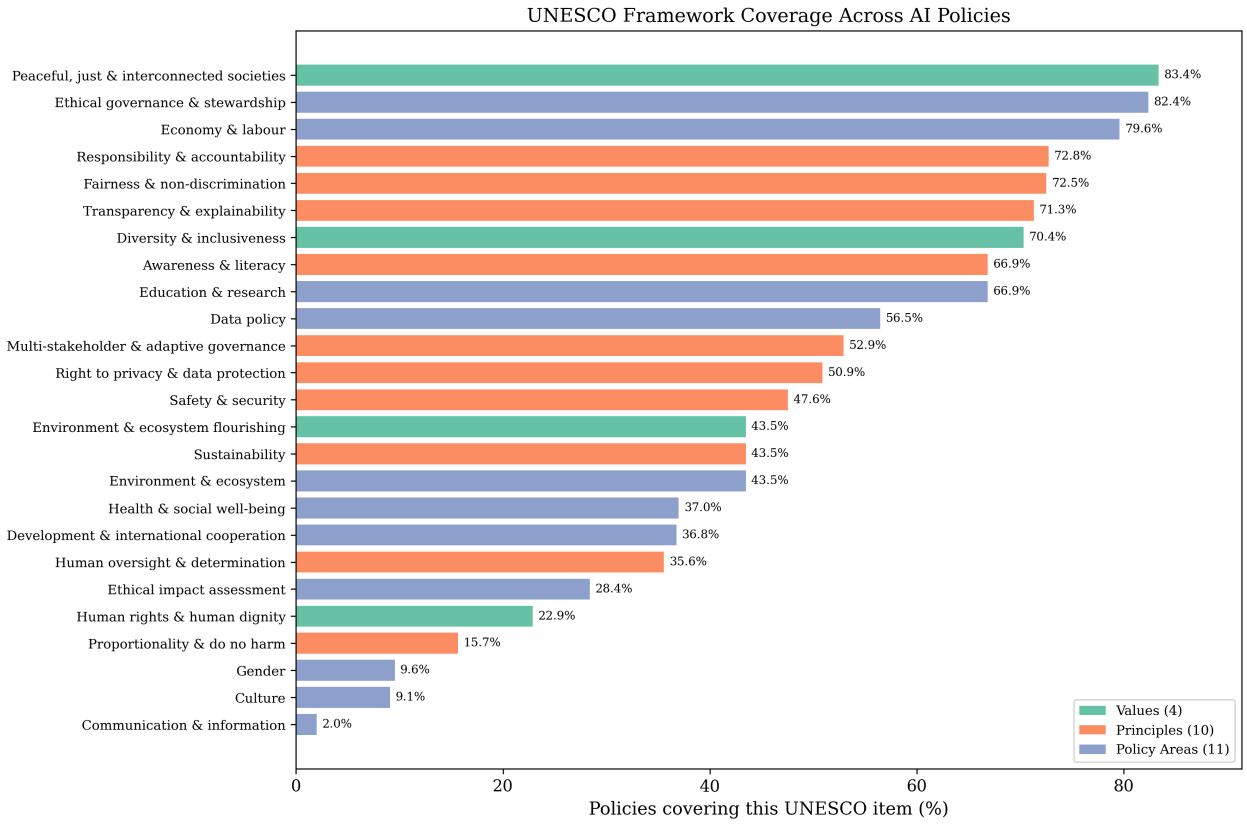

Figure 5.2: Coverage rates across all 25 UNESCO items, ordered by frequency. Five items exceed 70% coverage, while six fall below 30%.

The coverage landscape is highly uneven. The five most-covered items are:

Table 5.1: Most-addressed UNESCO items

UNESCO Item	Type	Coverage
Peaceful, just & interconnected societies	Value	83.4%
Ethical governance & stewardship	Policy area	82.4%
Economy & labour	Policy area	79.6%
Responsibility & accountability	Principle	72.8%
Fairness & non-discrimination	Principle	72.5%

By contrast, the five least-covered items are:

Table 5.2: Least-addressed UNESCO items

UNESCO Item	Type	Coverage
Communication & information	Policy area	2.0%
Culture	Policy area	9.1%
Gender	Policy area	9.6%
Proportionality & do no harm	Principle	15.7%
Human rights & human dignity	Value	22.9%

The **near-absence of communication and information** (2.0%) is particularly notable. The UNESCO Recommendation explicitly calls for policies to address AI's impact on media, information ecosystems, and freedom of expression—yet virtually no national policy addresses this component. Meanwhile, **human rights and human dignity**, the normative anchor of the entire framework, appears in only **22.9%** of policies. For a framework that explicitly grounds AI ethics in human rights, this represents a striking omission.

5.1.3 The Implementation Gap: Values vs. Principles vs. Policy Areas

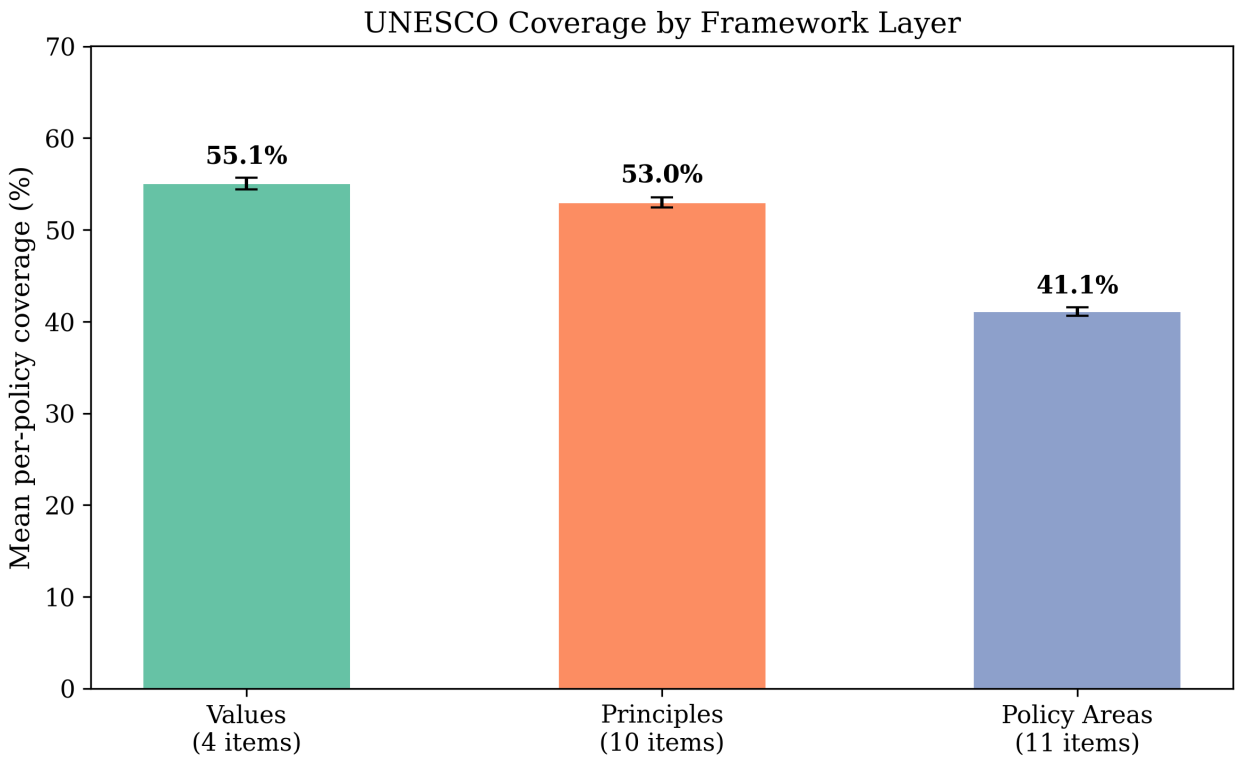

Figure 5.3: Mean coverage rates by UNESCO framework layer. Values and principles are better covered than policy action areas, revealing an "implementation gap."

The UNESCO Recommendation's three-tier structure reveals a telling gradient:

Table 5.3: Coverage by UNESCO framework layer

Layer	Mean Coverage	Items
Values	55.0%	4
Principles	53.0%	10
Policy action areas	41.1%	11

Values come easy; action areas require work. Policies readily declare broad ethical commitments but shy away from the concrete governance mechanisms needed to operationalize them. The 14-percentage-point gap between values and policy areas is the "principles-to-practice" problem that AI ethics critics have long identified—now documented at scale.

5.1.4 Coverage, Depth, and Gaps

Here is the uncomfortable finding: **coverage does not predict depth** ($r = 0.02$, $p = 0.94$). The most frequently invoked principles—transparency, accountability, fairness—tend to appear as brief rhetorical gestures, sometimes just a phrase, rather than deeply developed commitments.

Yet some rarely mentioned items receive serious treatment when they do appear. **Awareness and literacy** (66.9% coverage) averages **3.64** depth—paragraph-level engagement reflecting concrete education programmes. **Proportionality and do no harm**, covered by only 15.7% of policies, achieves **3.53** depth when present. The few policies that engage with it do so thoughtfully.

5.1.5 Depth Patterns

The depth heatmap shows that most UNESCO items, when mentioned, receive **sentence-level (3)** to **paragraph-level (4)** treatment. Very few items are engaged at section-level (5), suggesting that even the most substantive policies rarely dedicate entire sections to individual UNESCO items.

Depth by framework layer. Figure 5.6, Figure 5.7, and Figure 5.8 break the analysis down by UNESCO tier.

Across all three layers, the depth distributions are roughly similar, with most engagement occurring at depth levels 3–4. However, **policy action areas** show slightly higher depth when present — likely because they translate into concrete programmatic commitments (education, data governance, health) that inherently require more detailed treatment than abstract values.

The gaps. Figure 5.9 highlights the largest discrepancies between the Recommendation's aspirations and actual policy content.

The gap analysis identifies three categories of UNESCO alignment.

Well-integrated items (>70% coverage) include peaceful societies, ethical governance, economy and labour, responsibility, fairness, transparency, and diversity. These form the consensus core of global AI ethics—the items virtually all policy traditions address.

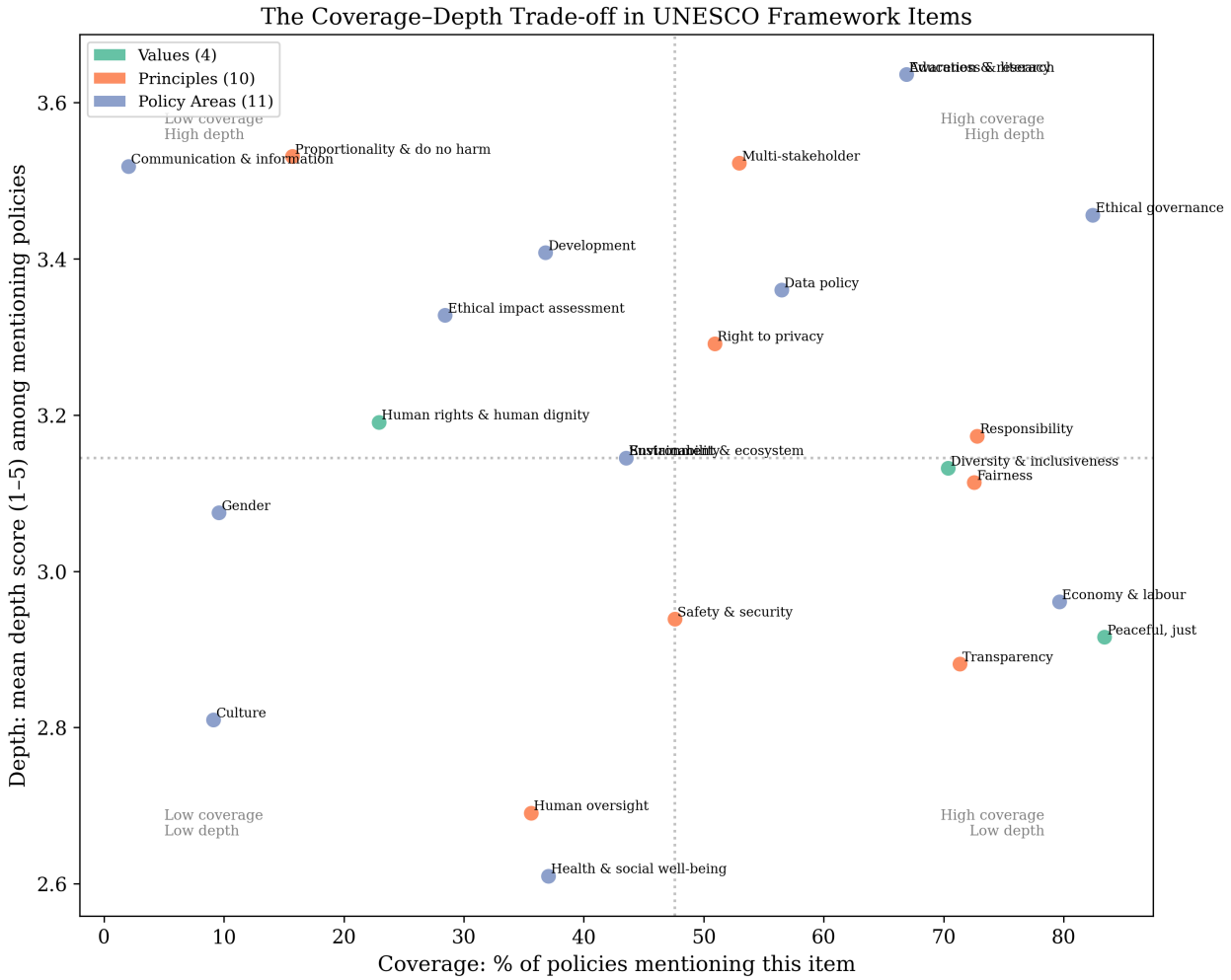

Figure 5.4: Scatter plot of coverage (% of policies mentioning an item) vs. mean depth (1–5 scale) for each UNESCO item. There is no significant correlation ($r = 0.02$, $p = 0.94$), indicating that breadth of mention does not predict substantive engagement.

Partially addressed items (30–70%) include data policy, education, privacy, safety, sustainability, environment, health, development cooperation, multi-stakeholder governance, and awareness. These appear in a majority or near-majority of policies but with significant variation.

Systematically neglected items (<30%) include human rights & dignity (22.9%), proportionality (15.7%), ethical impact assessment (28.4%), gender (9.6%), culture (9.1%), and communication & information (2.0%). These represent the most significant misalignment between the UNESCO Recommendation and actual global AI policy practice.

The pattern of neglect is not random; it reflects political economy. **Gender** (9.6%) requires governments to address algorithmic bias, gendered data gaps, and the differential impacts of automation on women's employment—issues that create regulatory costs and may conflict with industry preferences for minimal governance. **Culture** (9.1%) requires engagement with indigenous knowledge systems, linguistic diversity, and the rights of cultural minorities—domains where AI governance intersects

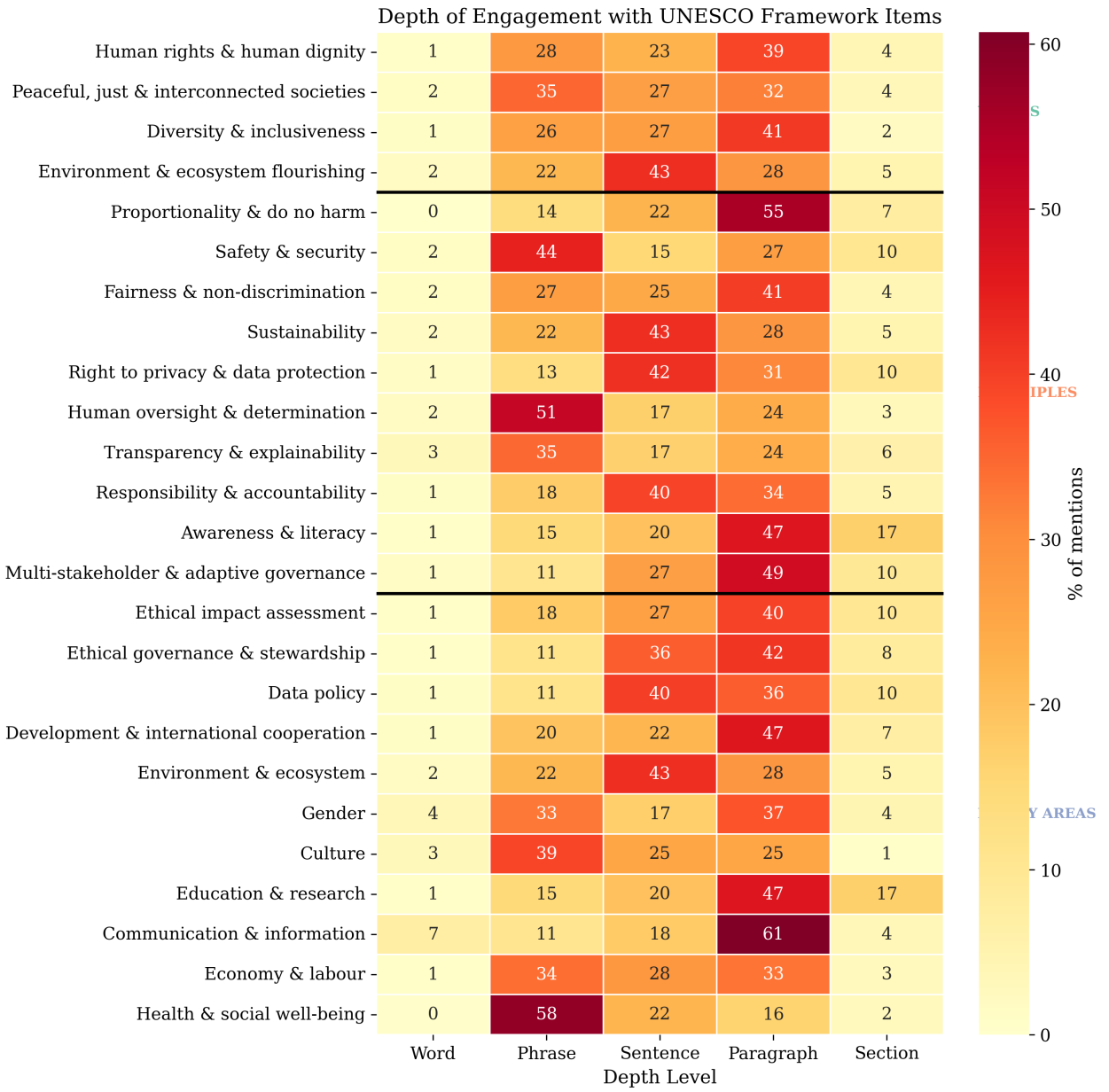

Figure 5.5: Depth heatmap across all 25 UNESCO items, showing the distribution of engagement levels from word (1) to section (5). Most engagement occurs at sentence (3) to paragraph (4) level.

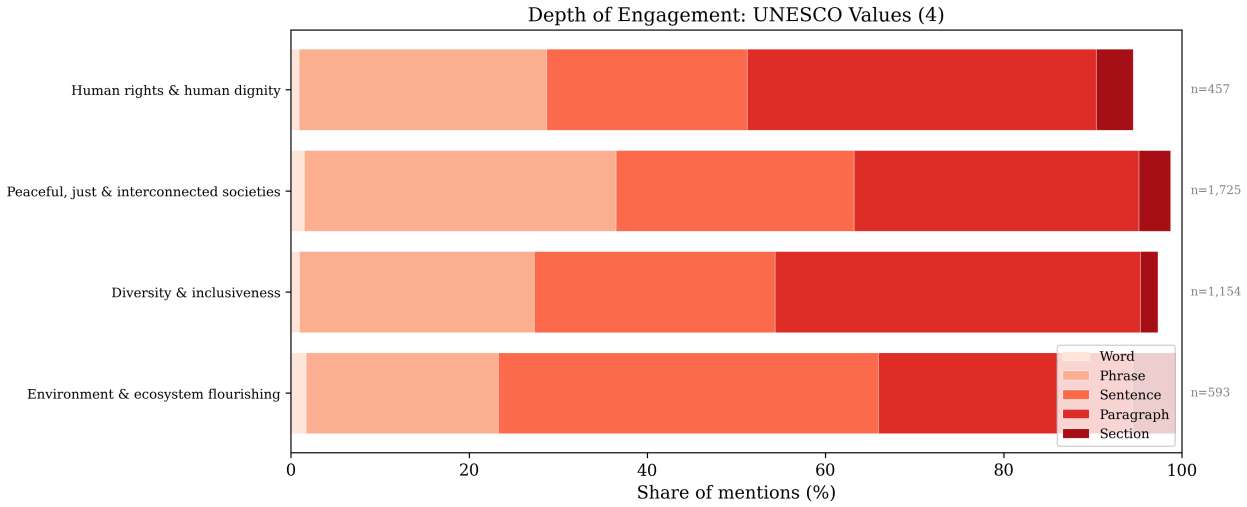

Figure 5.6: Depth distributions for the 4 UNESCO values.

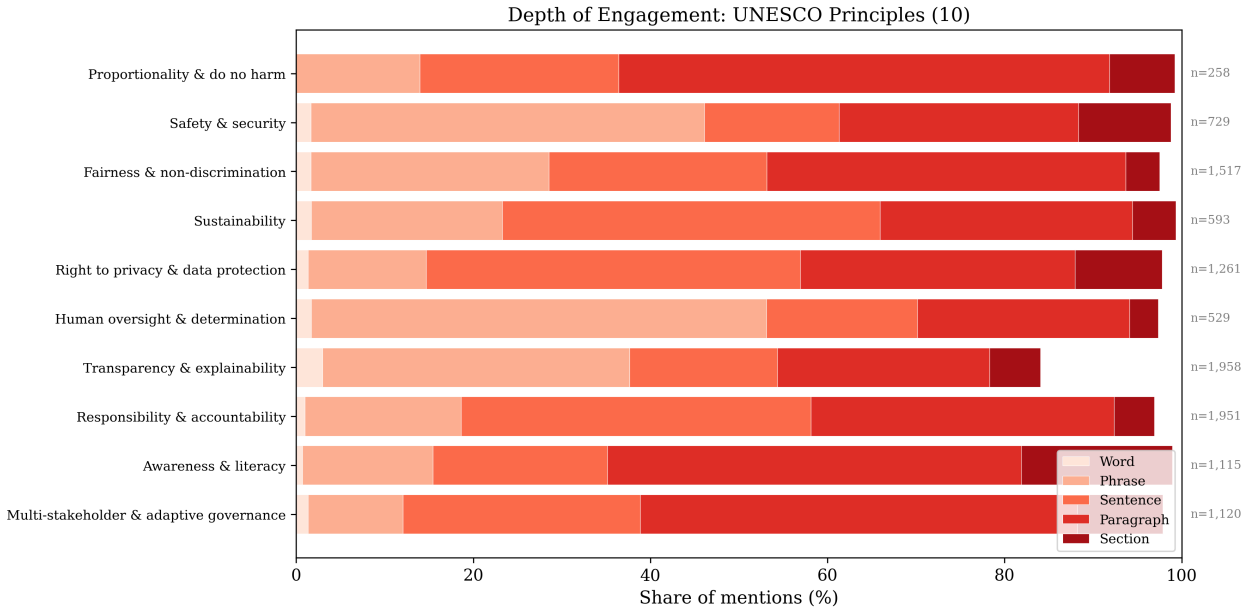

Figure 5.7: Depth distributions for the 10 UNESCO principles.

with politically sensitive identity politics. **Communication & information** (2.0%) addresses AI's impact on media, disinformation, and freedom of expression—a domain where governments may resist governance constraints on their own information management capabilities. **Human rights & dignity** (22.9%) is perhaps the most revealing absence: for a framework that explicitly grounds AI ethics in human rights, this gap confirms that most countries approach AI governance as technology management rather than rights protection.

The neglect of human rights as a foundational frame, and the near-absence of gender-specific and cultural considerations, suggests that the global AI policy landscape remains shaped by a technology-

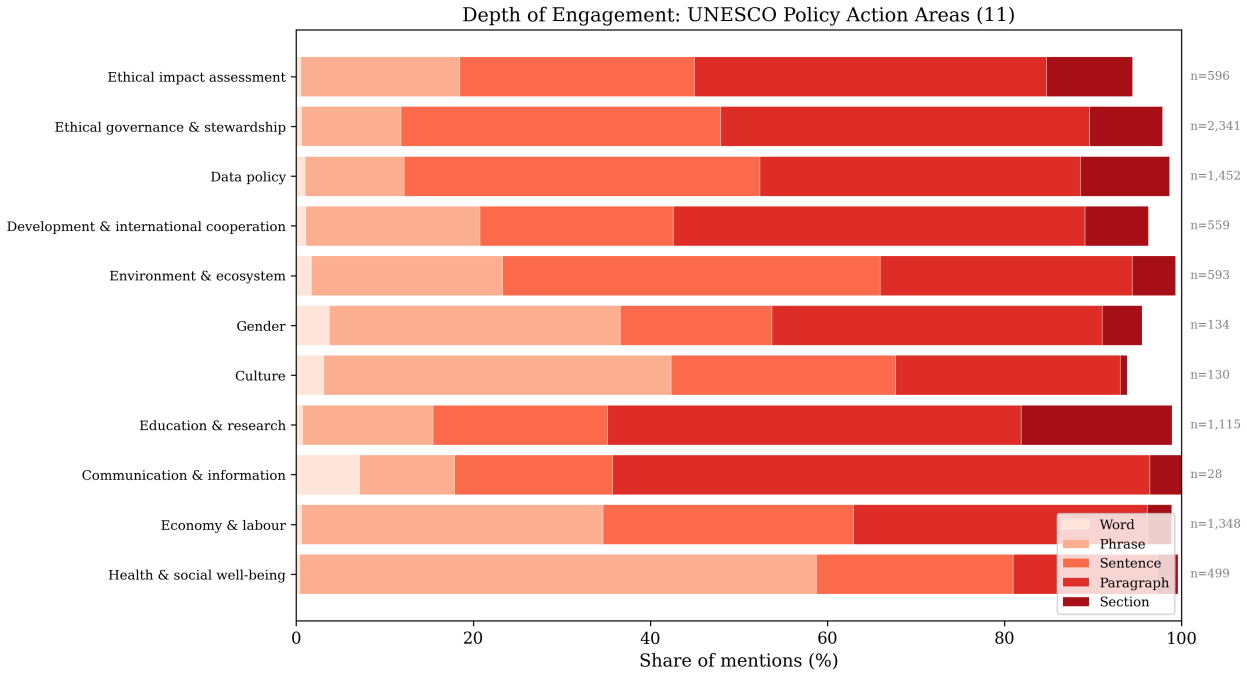

Figure 5.8: Depth distributions for the 11 UNESCO policy action areas.

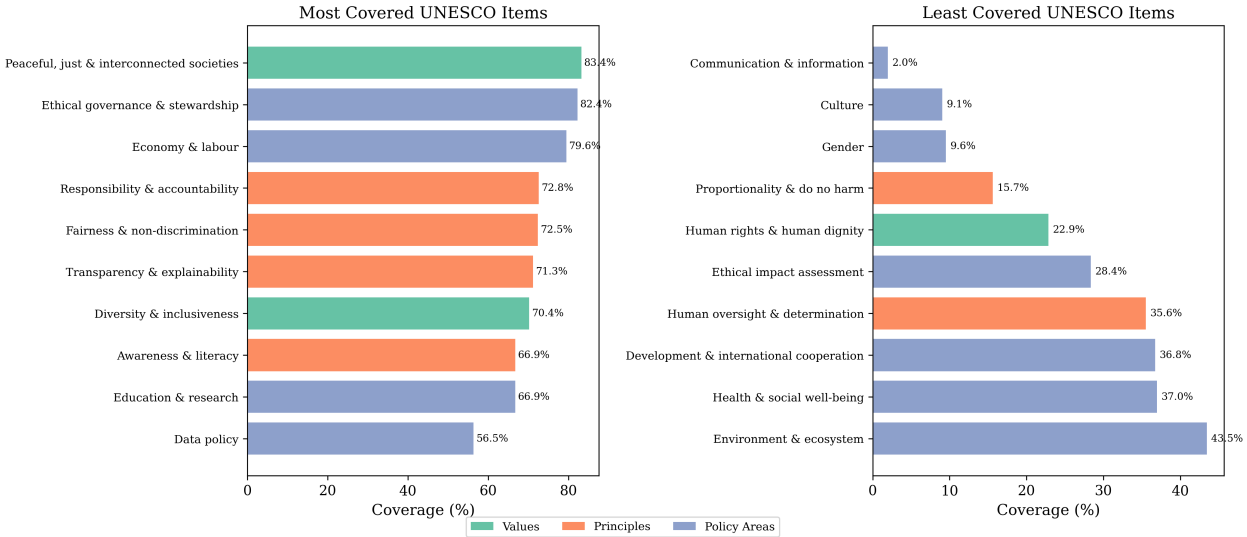

Figure 5.9: Top and bottom UNESCO items by coverage, highlighting the largest gaps between the Recommendation's aspirations and actual policy content.

centric governance paradigm rather than the rights-based approach UNESCO advocates. The items that *are* well-integrated—transparency, fairness, accountability—are precisely those that align with existing regulatory traditions and impose the least institutional disruption.

6 UNESCO Alignment Determinants

6.1 What Drives UNESCO Alignment?

The conventional wisdom in AI governance discourse holds that wealthier countries should demonstrate stronger policy alignment. Richer countries possess greater institutional capacity, regulatory expertise, and resources for comprehensive policy development. The analysis reveals a different pattern.

6.1.1 The Income Divide That Isn't

Table 6.1: Income-group comparison for UNESCO alignment

Metric	Value
High income mean (N = 1,049)	53.5
Developing mean (N = 204)	53.5
Welch's t	0.013
p-value	0.99
Cohen's d	**0.001**

The income gap is **effectively zero** ($d = 0.001$). High-income and developing countries demonstrate identical means and distribution.

The UNESCO framework, negotiated with input from all member states, may function as a normative template equally accessible regardless of national wealth. It specifies *what* to address rather than *how* to address it—and the "what" requires no financial resources to adopt. Capacity and ethics scores demonstrated small but significant income gaps ($d = 0.30$ and $d = 0.20$); UNESCO alignment demonstrates none.

Item-level income gaps. Figure 6.2 and Figure 6.3 reveal where aggregate parity masks item-level heterogeneity.

The aggregate null result conceals meaningful item-level heterogeneity. Two items show **statistically significant** income gaps — both favouring developing countries:

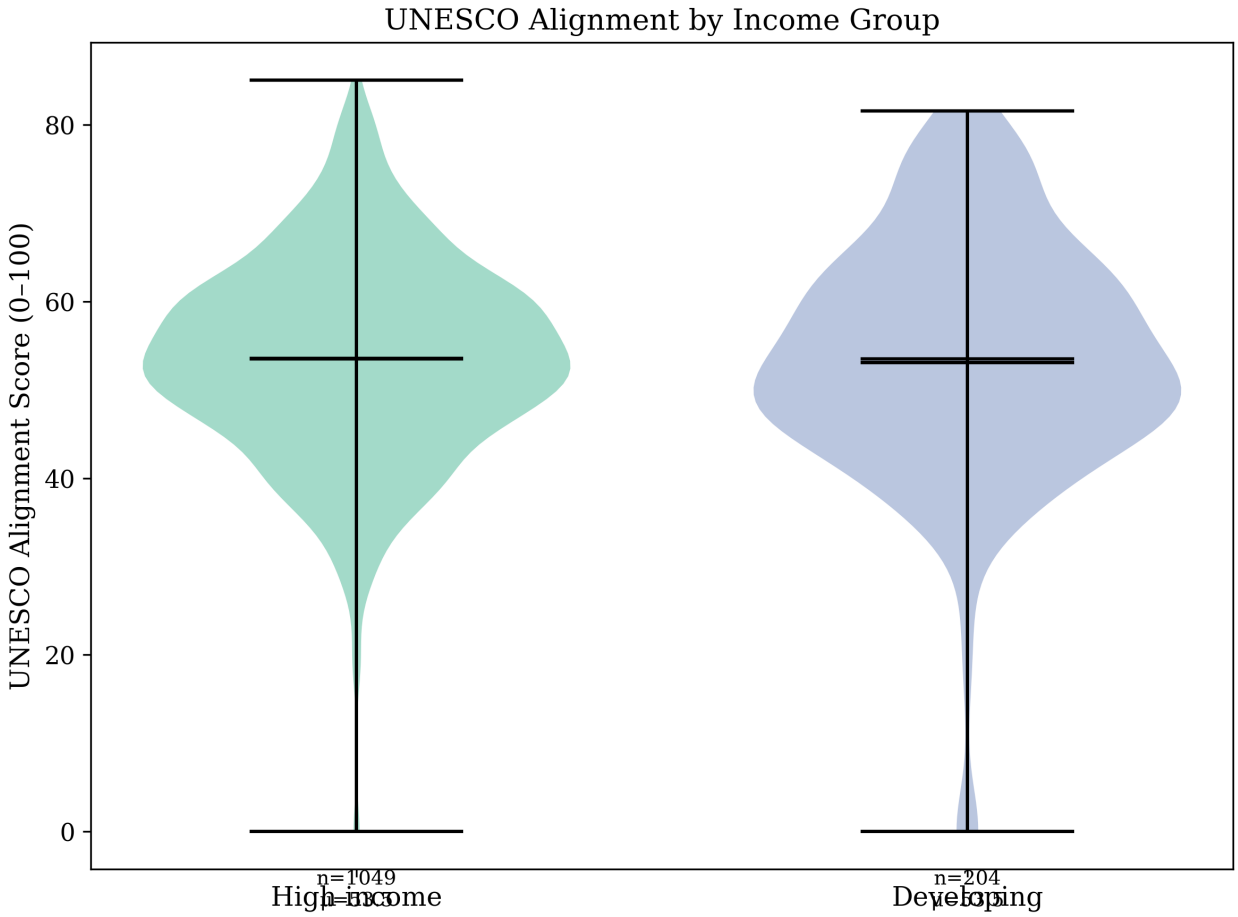

Figure 6.1: Violin plots of UNESCO alignment scores by income group. The distributions are remarkably similar, with near-complete overlap.

Table 6.2: Significant item-level income gaps

UNESCO Item	HI Coverage	Dev Coverage	Gap (pp)	p
Health & social well-being	34.1%	53.4%	**−19.3**	$< .001$
Gender	7.5%	13.7%	**−6.2**	.006

Developing countries are nearly **20 percentage points more likely** to address health and social well-being. Two factors explain this: first, health-sector AI deployment is particularly salient in low- and middle-income countries, where AI is framed primarily as a development tool rather than a general-purpose technology; second, international development frameworks (SDGs, WHO guidelines) consistently emphasize health equity, and developing-country policies reflect that influence.

The gender gap (+6.2pp for developing countries) tells a similar story. UN system frameworks foreground gender mainstreaming, and developing-country policies are more responsive to that emphasis than their high-income counterparts.

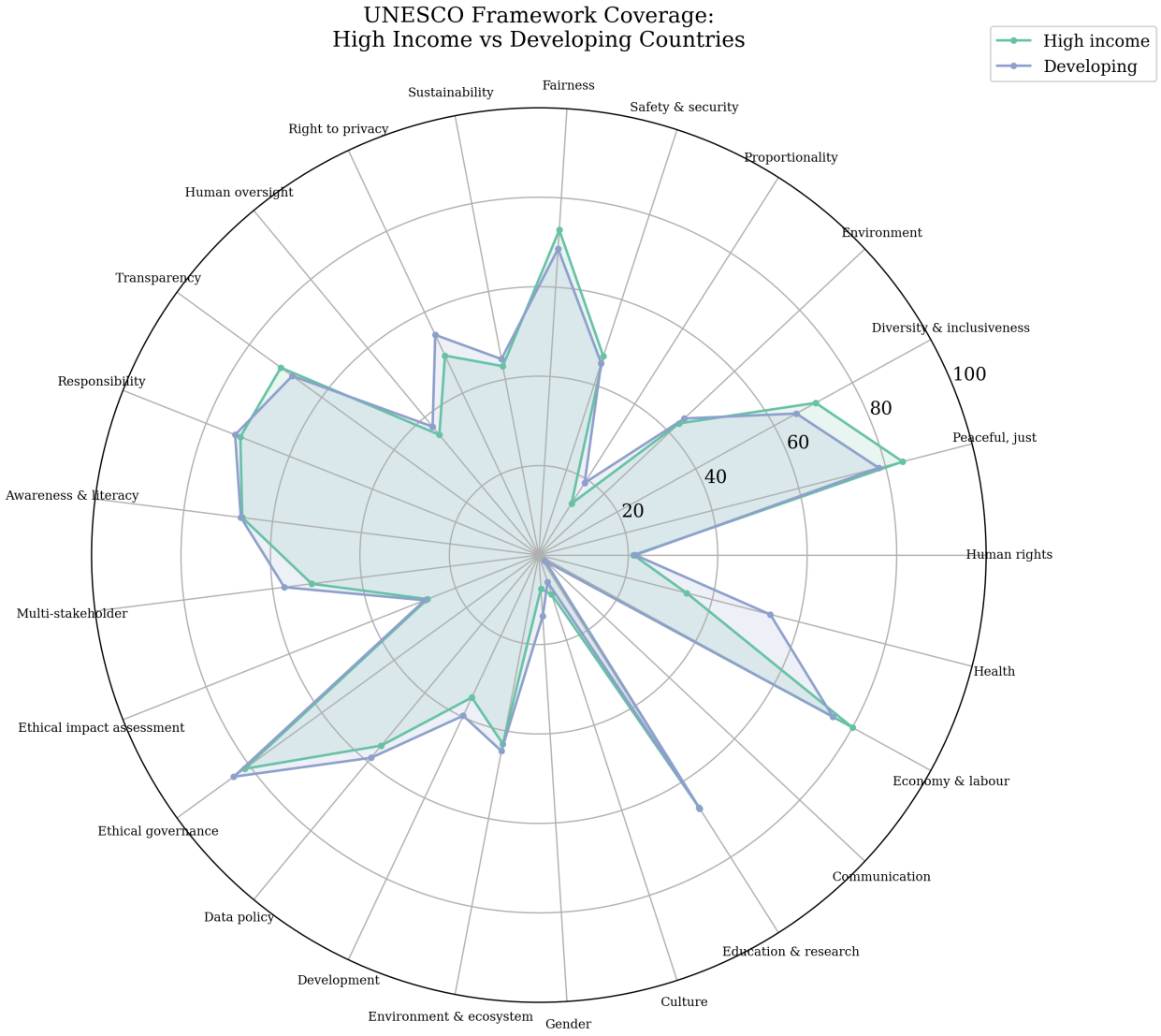

Figure 6.2: Radar plot comparing UNESCO item coverage between high-income and developing countries. The profiles are largely overlapping, with notable divergences on health and gender.

6.1.2 Regional Patterns

Regional variation in UNESCO alignment is present but modest. All regions fall within the 45–60 range on the 0–100 scale. The regional heatmap reveals that while overall alignment levels are similar, regional profiles diverge on *which* UNESCO items get attention.

European policies emphasise transparency, accountability, and privacy—the GDPR tradition writ large. The EU AI Act and national transpositions in France, Germany, and the Netherlands embed algorithmic transparency and accountability into binding regulation, producing high scores on these components while largely ignoring environment and gender. This is consistent with Brad-

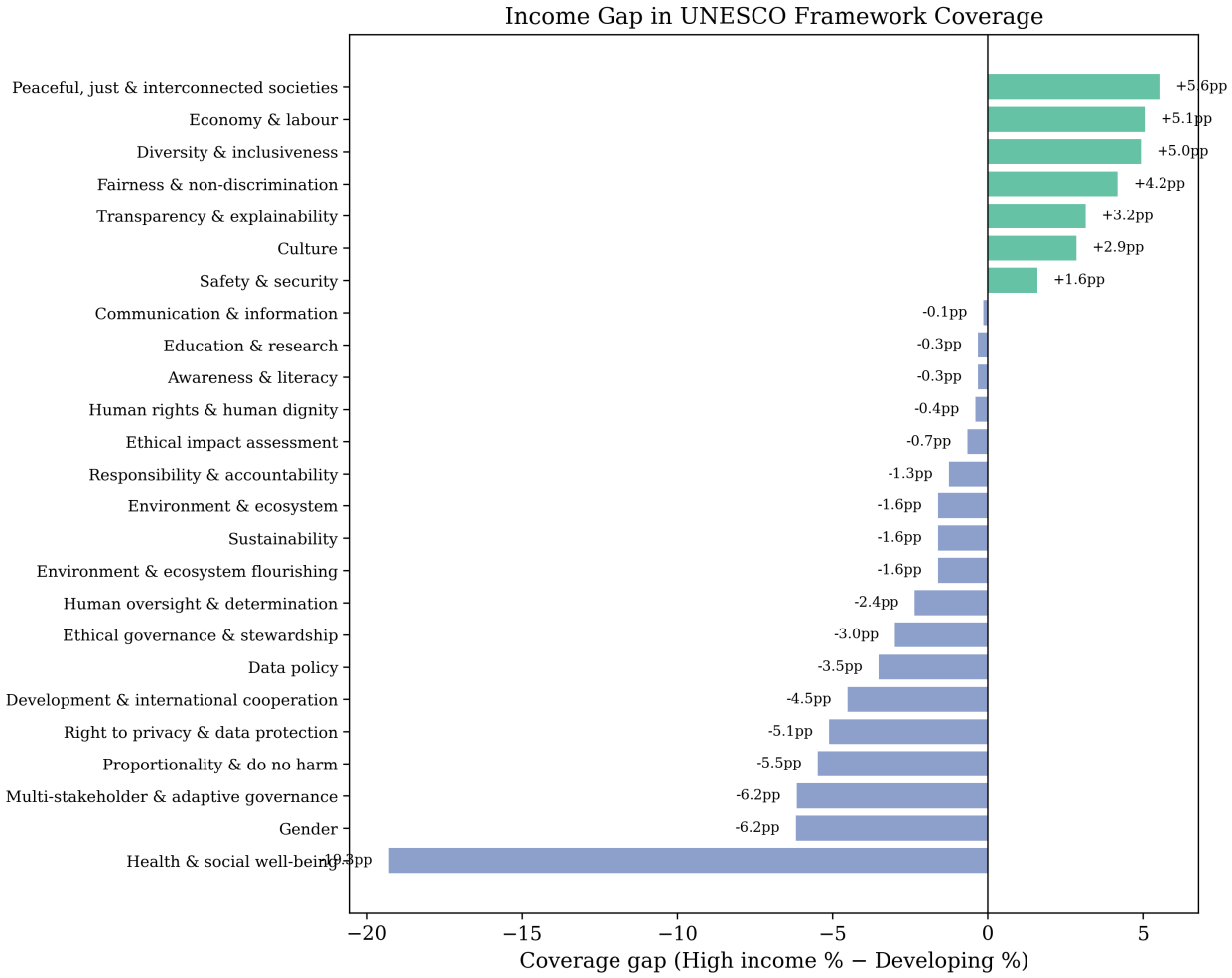

Figure 6.3: Bar chart of UNESCO item coverage gaps between income groups, highlighting statistically significant differences.

ford (2020)'s Brussels Effect: European regulatory standards diffuse internationally through market mechanisms, but the diffusion is selective—technical requirements travel more easily than values-based commitments.

African policies show the highest overall alignment (mean 1.78), engaging more broadly with development cooperation, multi-stakeholder governance, and sustainability. Countries like Kenya, Rwanda, and Mauritius have adopted comprehensive national AI strategies that reference the UNESCO framework explicitly, using it as a governance template where domestic regulatory traditions are thinnest. This pattern is consistent with Acharya (2004)'s prediction that new governance frameworks gain traction where they fill institutional vacuums rather than competing with established arrangements.

Asian policies display the widest internal variation. Japan and South Korea emphasise human-centred AI and data governance, reflecting established technology policy traditions. Singapore's Model AI Governance Framework prioritises practical implementation guidance over normative

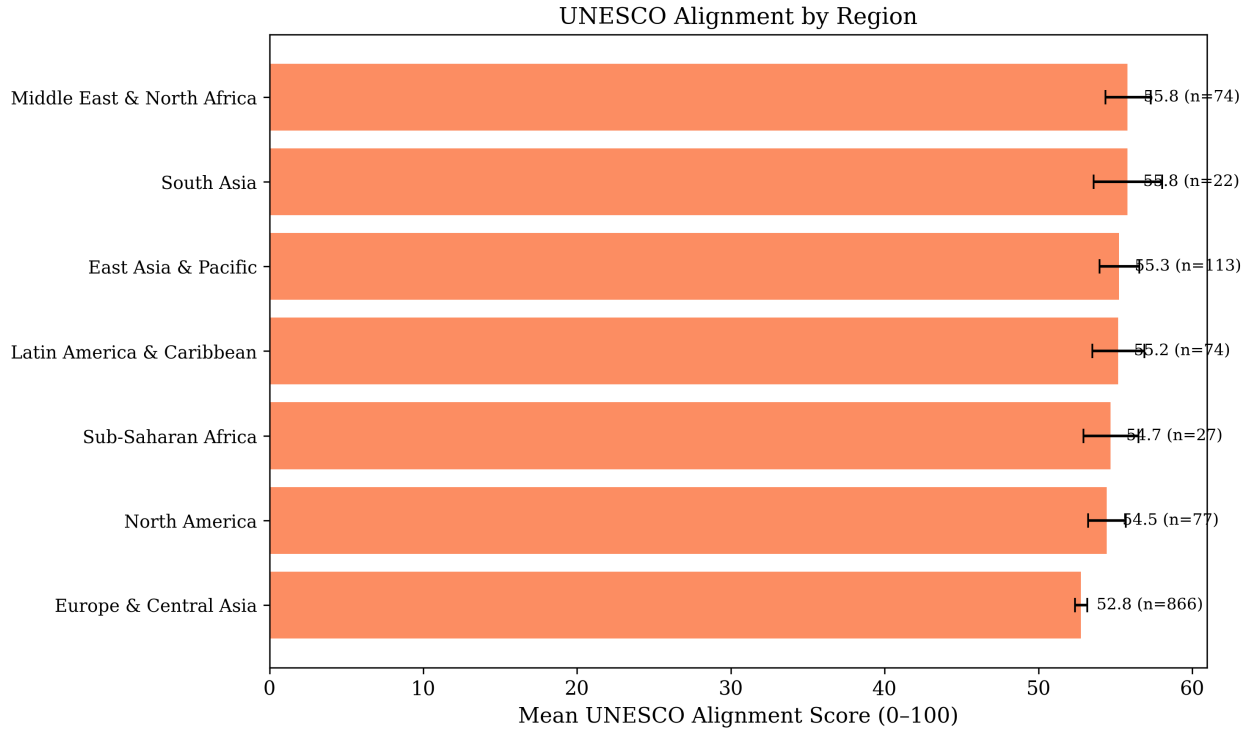

Figure 6.4: Mean UNESCO alignment scores by region. Regional variation is modest, with most regions clustered within a 10-point range.

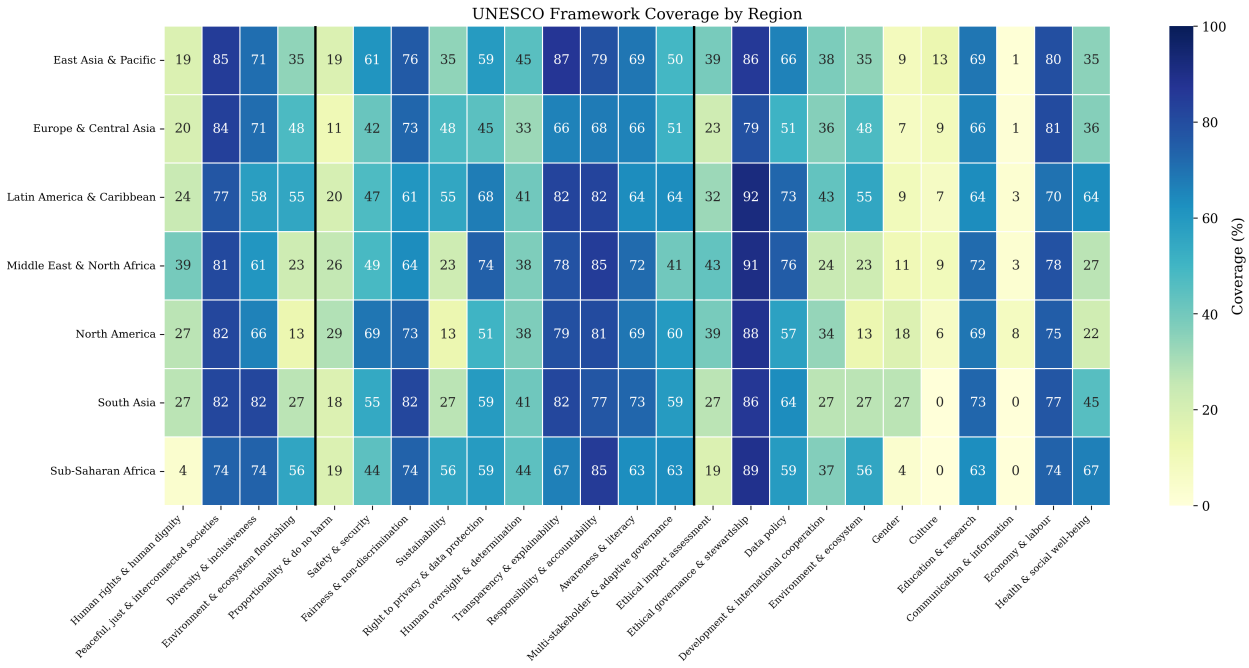

Figure 6.5: Regional heatmap showing UNESCO item coverage rates by region. Each row is a region, each column a UNESCO item.

breadth, producing moderate alignment scores despite sophisticated governance. China's AI governance approach emphasises safety and security alongside state capacity, engaging selectively with UNESCO components that align with domestic priorities.

North American policies cluster around safety, security, and economy, reflecting a market-oriented governance philosophy. The United States and Canada both score strongly on accountability and risk management but weakly on development cooperation and cultural considerations—consistent with a regulatory approach that prioritises innovation-friendly governance over redistributive or rights-based frameworks.

6.1.3 Policy Characteristics and Regression Analysis

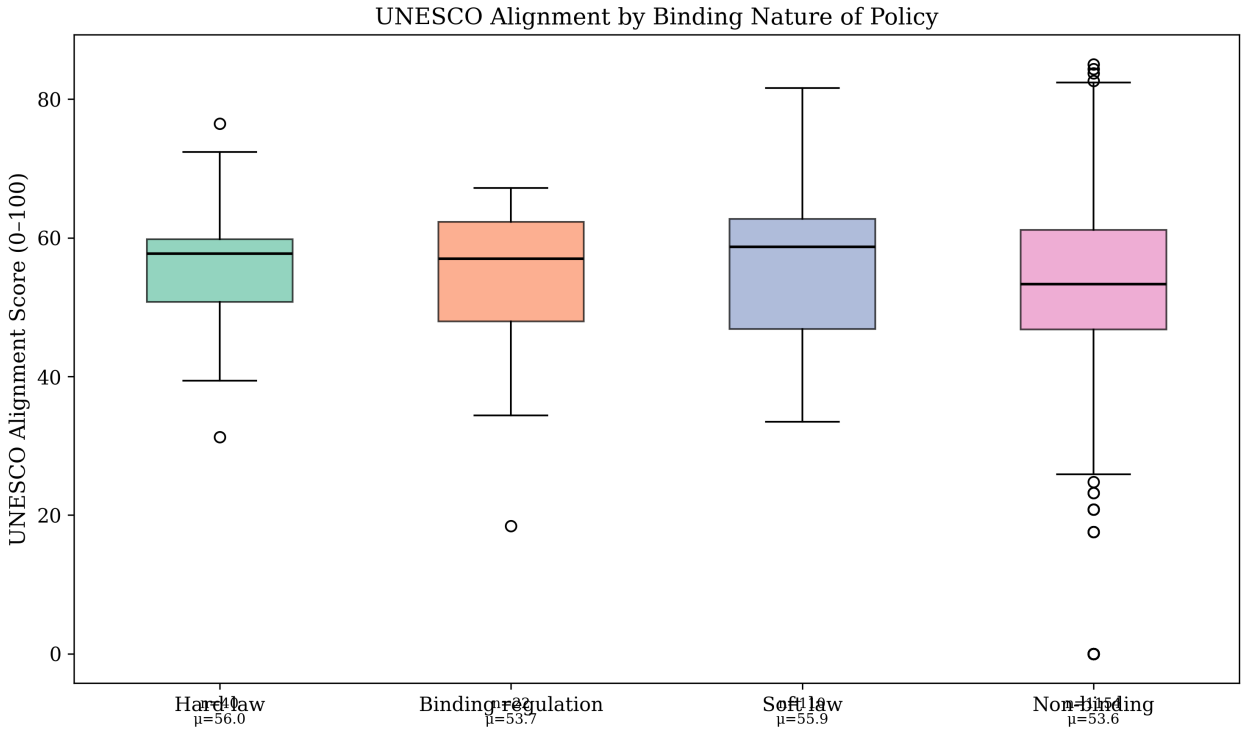

Figure 6.6: UNESCO alignment scores by binding nature. Differences are small and not statistically significant ($F = 1.66$, $p = 0.17$).

Table 6.3: UNESCO alignment by binding nature

Binding Nature	N	Mean	SD	Coverage
Hard law	40	56.0	9.2	47.3%
Soft law	110	55.9	10.7	50.1%
Binding regulation	22	53.7	11.9	41.5%
Non-binding	1,154	53.6	12.4	48.0%

The ANOVA is not significant ($F = 1.66$, $p = 0.17$), indicating that binding nature does not strongly predict UNESCO alignment. However, a suggestive pattern emerges: hard law and soft law instruments score 2–3 points higher than non-binding documents. This small difference may reflect the greater drafting effort and stakeholder consultation that formal legal instruments typically undergo.

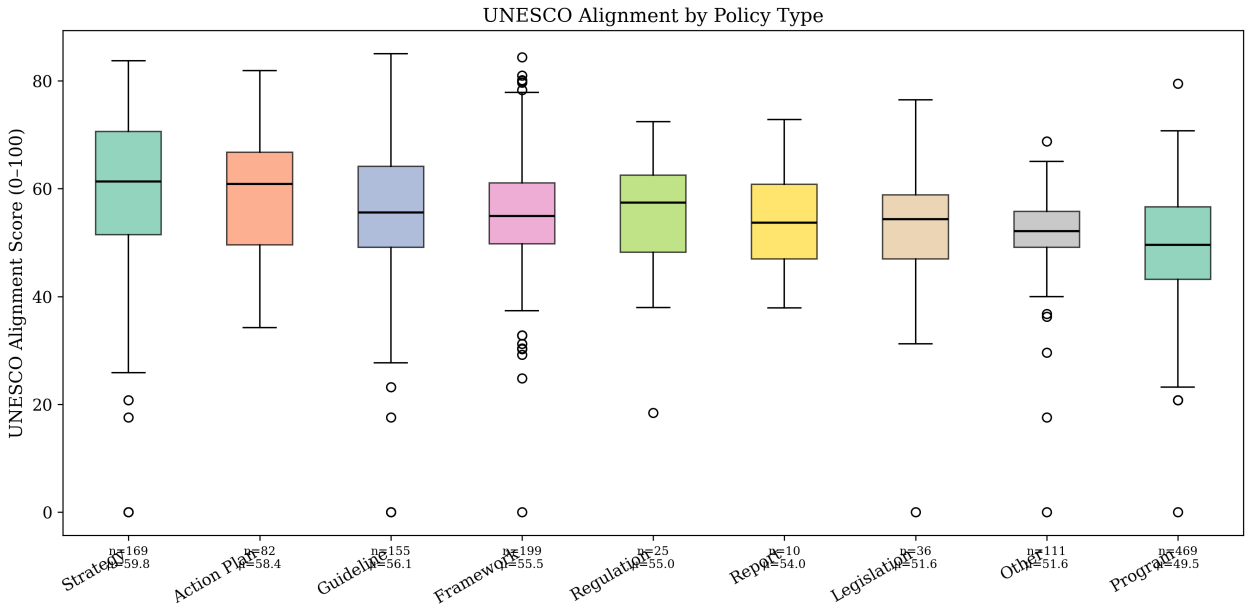

Figure 6.7: UNESCO alignment scores by policy type. Variation across policy types is substantial, with strategies and action plans tending to score highest.

Policy type shows more variation than binding nature. National AI strategies and action plans, which are typically comprehensive, forward-looking documents, tend to achieve higher UNESCO alignment than narrower regulatory instruments or sectoral guidelines.

Multivariate regression. To disentangle the relative contributions of structural factors, we estimate three OLS models with heteroskedasticity-consistent (HC1) standard errors. The dependent variable is the UNESCO alignment score (0–100).

Model 1: Structural covariates only.

Table 6.4: Regression Model 1: Structural predictors only ($N = 1{,}253$, $R^2 = 0.008$)

Variable	β	SE	p
Intercept	−102.4	153.7	.505
Developing country	0.49	1.67	.769
Post-UNESCO era	**−1.89**	0.79	**.017**
Log GDP per capita	0.42	0.92	.647
Hard law	0.32	1.42	.821
Soft law	**2.65**	1.08	**.014**
Year	0.08	0.08	.321

Model 1 has very low explanatory power ($R^2 = 0.008$), confirming that structural country-level characteristics explain almost none of the variance in UNESCO alignment. Two variables are significant: the **post-UNESCO era** dummy ($\beta = -1.89$, $p = .017$) — alignment is slightly *lower* in the post-2021 period — and **soft law** ($\beta = 2.65$, $p = .014$) — soft law instruments score modestly higher.

The negative post-UNESCO coefficient is counterintuitive and is explored further in Section 8.1. It likely reflects compositional change in the policy corpus rather than declining alignment.

Model 2: Adding capacity and ethics scores.

Table 6.5: Regression Model 2: With capacity and ethics scores ($N = 1{,}253$, $R^2 = 0.348$)

Variable	β	SE	p
Developing country	−0.17	1.38	.900
Post-UNESCO era	**−1.40**	0.69	**.043**
Log GDP per capita	−0.47	0.74	.525
Hard law	**−9.31**	1.39	**< .001**
Soft law	**−4.22**	0.91	**< .001**
Year	−0.09	0.09	.305
Capacity score	**3.22**	0.47	**< .001**
Ethics score	**11.55**	0.63	**< .001**

Adding capacity and ethics scores dramatically increases explanatory power ($R^2 = 0.348$). The two strongest predictors:

Ethics score ($\beta = 11.5$, $p < .001$): policies with richer ethical content align more closely with UNESCO's fundamentally ethics-focused framework.

Capacity score ($\beta = 3.2$, $p < .001$): implementation capacity also predicts alignment, though less strongly.

Binding nature coefficients reverse sign once content is controlled. Hard law ($\beta = -9.3$) and soft law ($\beta = -4.2$) now show *lower* alignment than non-binding documents—formally binding instruments are narrower in scope and cover fewer UNESCO items, even if they engage more deeply on the items they address.

Model 3: Income × Post-UNESCO interaction. The interaction between developing-country status and the post-UNESCO era is not significant ($\beta = -1.77$, $p = .393$), indicating that the (slight) post-UNESCO decline in alignment is not differentially driven by developing countries. The UNESCO Recommendation's adoption does not appear to have produced a differential effect by income group.

Summary. UNESCO alignment is a policy-level phenomenon, not a country-level one. National income, GDP per capita, and region explain almost none of the variance. What matters is policy content: the depth of ethics and capacity provisions. Alignment with international frameworks depends not on wealth but on ambition and comprehensiveness.

7 UNESCO Alignment Clusters

7.1 Alignment Archetypes: A Cluster Analysis

Not all policies engage with UNESCO the same way. Some adopt broadly; others pick specific priorities; many do the minimum. Cluster analysis reveals four distinct archetypes—and income has nothing to do with which cluster a policy falls into.

7.1.1 Clustering Methodology

K-means clustering was applied to the 25-dimensional binary coverage vector for each policy (1 = UNESCO item mentioned, 0 = not mentioned). To determine the optimal number of clusters, silhouette scores were evaluated for $k = 3, 4, 5, 6$:

Table 7.1: Silhouette scores for candidate cluster solutions

k	Silhouette Score
3	0.192
4	**0.204**
5	0.199
6	0.190

The 4-cluster solution maximises the silhouette score ($s = 0.204$) and yields substantively interpretable archetypes. While the silhouette values are modest (reflecting the inherent overlap in policy coverage patterns), the resulting clusters display clearly differentiated profiles.

7.1.2 The Four Archetypes

Table 7.2: Cluster profiles: UNESCO alignment archetypes

Cluster	N	Mean Alignment	Coverage	HI %	Dev %
Comprehensive aligners	365	61.4	58.8%	80%	14%
Moderate aligners	337	60.1	55.1%	79%	14%
Selective aligners	204	52.9	52.1%	75%	20%
Minimal engagement	420	42.8	31.2%	81%	16%

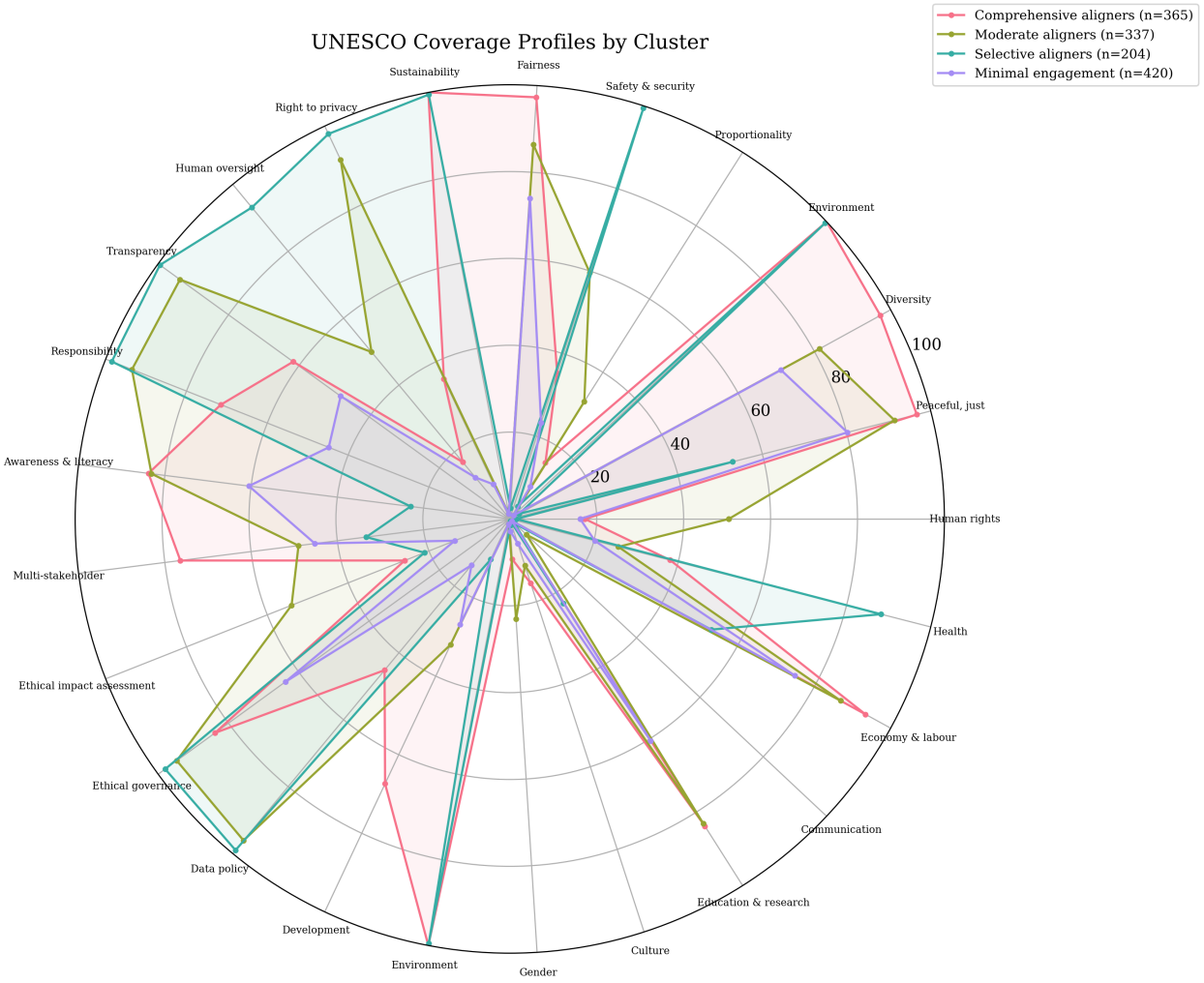

Figure 7.1: Radar charts showing the UNESCO item coverage profile of each cluster. The four archetypes display clearly differentiated patterns of engagement with the 25 UNESCO items.

Archetype 1: Comprehensive aligners (n = 365). These policies achieve the highest mean alignment (61.4) and broadest coverage (58.8%). They show near-universal engagement with environment/sustainability (100%), diversity (97%), peaceful societies (97%), and fairness (97%); strong coverage of awareness and literacy (84%), ethical governance (84%), and multi-stakeholder governance (76%); but weaker attention to safety (35%), privacy (36%), and human oversight (17%)—a values-first orientation.

This archetype represents the "UNESCO ideal": policies that broadly embrace the Recommendation's normative vision, typically national AI strategies taking a comprehensive, whole-of-government approach. Exemplars include Kenya's National AI Strategy and Canada's Pan-Canadian AI Strategy, both of which address the full spectrum of UNESCO values, sustainability, and multi-stakeholder governance from the outset.

Archetype 2: Moderate aligners (n = 337). With mean alignment of 60.1 (close to the comprehensive cluster), these policies achieve similar scores through a different profile: strong on technical principles (transparency 94%, responsibility 94%, privacy 91%, fairness 86%, human oversight 50%), strong on data governance (96%), but weak on environment/sustainability (1.2%) and gender (23%)—a notable blind spot.

This archetype emphasises the regulatory-technical dimension of AI ethics: data protection, algorithmic transparency, and accountability mechanisms. It reflects the European regulatory tradition. Policies from Germany, France, and the Netherlands cluster here, as do Singapore's Model AI Governance Framework and Japan's Social Principles of Human-Centric AI—all instruments that excel on technical governance while giving limited attention to environmental and gender dimensions.

Archetype 3: Selective aligners (n = 204). This cluster ($\mu = 52.9$) is defined by stark specialisation: near-universal engagement with environment (100%), safety (100%), privacy (98%), data policy (99%), transparency (100%), responsibility (99%), and human oversight (93%); near-zero engagement with diversity (2.5%), fairness (2.5%), culture (0.5%), and gender (1.5%); and the highest health coverage of any cluster (88.2%).

This is the technocratic safety archetype: policies focused on risk management, data protection, and sectoral safety while largely ignoring broader social and cultural dimensions. The higher proportion of developing countries (20%) may reflect the influence of sector-specific AI regulation in health and environment. Typical examples include sectoral AI guidelines from Brazil's health ministry and India's NITI Aayog sectoral papers, which address safety, data protection, and domain-specific risks in depth but do not engage with the broader normative architecture UNESCO envisions.

Archetype 4: Minimal engagement (n = 420). The largest cluster ($\mu = 42.8$, coverage = 31.2%) shows limited UNESCO engagement. Coverage exceeds 50% on only three items (peaceful societies 80%, economy 75%, fairness 74%). Most items are addressed by fewer than half of these policies, with near-zero attention to environment (1.2%), sustainability (1.2%), gender (2.9%), and culture (6.0%).

These are narrower instruments—sectoral guidelines, technical standards, early-stage consultations—that address specific governance concerns without engaging the UNESCO framework's full breadth. Examples include early-stage consultation documents from smaller jurisdictions and narrow technical standards focused on a single domain (e.g., autonomous vehicles, financial services) without broader normative framing.

7.1.3 Geographic and Income Patterns

The geographic distribution shows several patterns. European jurisdictions are distributed across comprehensive and moderate clusters, reflecting the continent's dual emphasis on rights-based governance and technical regulation. North American policies lean toward moderate and selective archetypes, consistent with a market- and safety-oriented approach. African and Asian policies appear across all clusters—regional governance traditions are less deterministic than often assumed.

Income composition. Figure 7.3 examines whether cluster membership tracks national wealth.

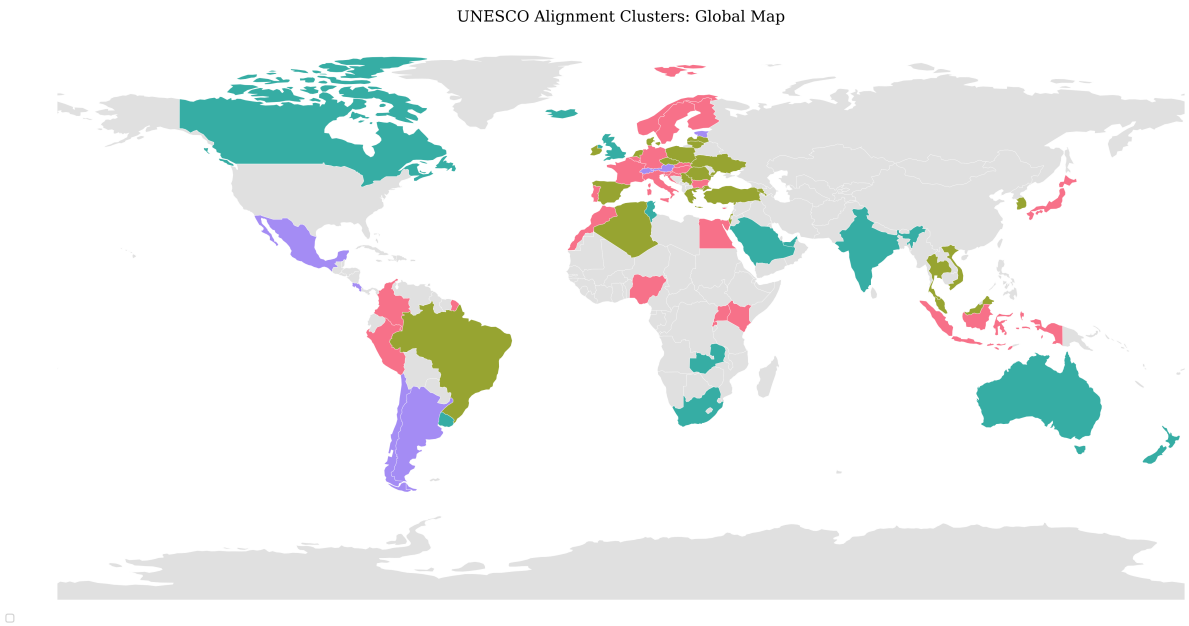

Figure 7.2: World map coloured by the most common UNESCO alignment cluster for each jurisdiction. The map reveals geographic clustering of alignment archetypes.

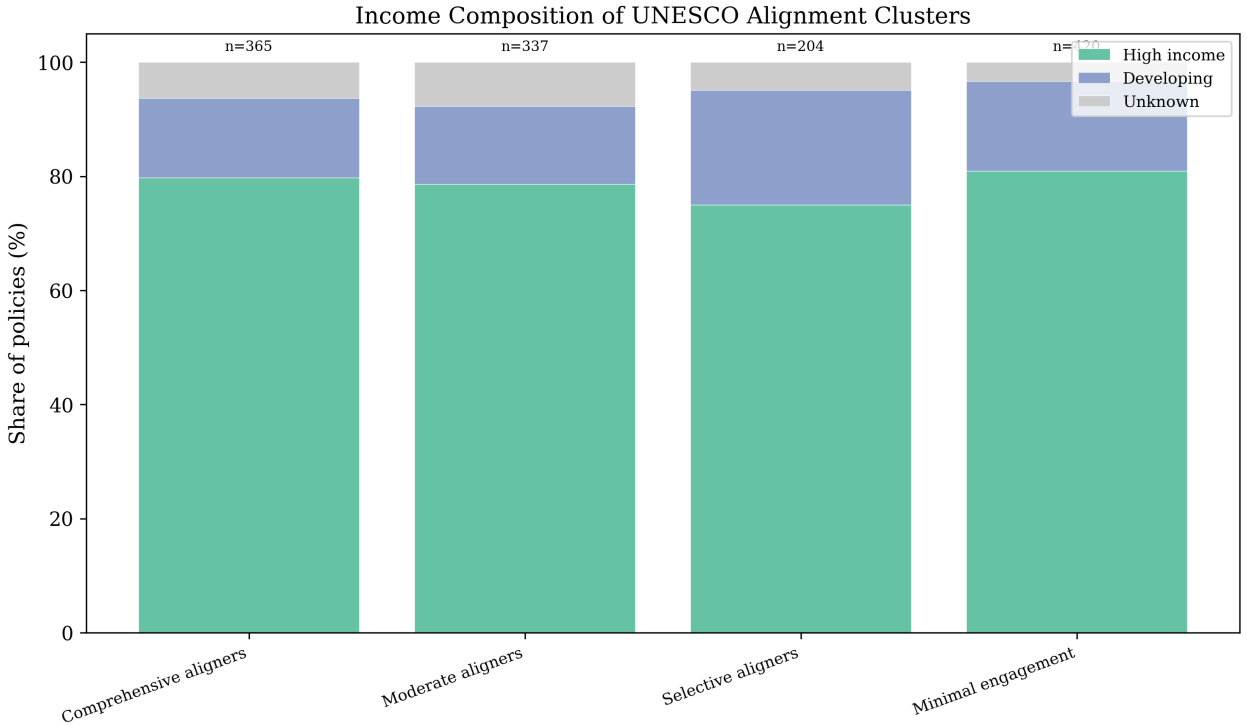

Figure 7.3: Income composition of each UNESCO alignment cluster. The proportion of high-income vs. developing countries is remarkably similar across clusters.

Income composition is nearly identical across all four clusters. High-income countries constitute 75–81% of each; developing countries 14–20%. The chi-squared test for independence between cluster membership and income group is not significant.

This confirms the central finding from Section 6.1.1: UNESCO alignment reflects policy design choices, not national wealth. A developing country is no more likely to fall into the "minimal engagement" cluster than a high-income country. What matters is whether a document takes a comprehensive values-based approach, a narrow regulatory focus, or a selective sectoral lens—not where it originates.

Interpreting the archetypes. The four-cluster solution maps onto recognisable governance traditions in AI policy:

Table 7.3: Cluster interpretation framework

Archetype	Governance Tradition	UNESCO Engagement
Comprehensive	Values-based, whole-of-government	Broad and deep
Moderate	Regulatory-technical (EU-influenced)	Strong on principles, weak on environment
Selective	Technocratic-safety	Deep but narrow
Minimal	Sectoral or early-stage	Limited across all dimensions

These archetypes are not normatively ranked. A "selective aligner" policy may be highly effective within its domain even if it does not address all 25 UNESCO items. The cluster analysis is descriptive: it shows *how* the global policy landscape engages with the UNESCO framework, not *how well* individual policies achieve their governance objectives.

Nonetheless, the existence of a large "minimal engagement" cluster (n = 420, 32% of all policies) suggests that a substantial share of the global AI policy corpus either predates the UNESCO Recommendation or does not engage with the kind of comprehensive ethical framework it envisions. Helping these policies broaden their normative scope is perhaps the most actionable implication of this analysis.

8 UNESCO Alignment Dynamics

8.1 Temporal Dynamics: Before and After UNESCO

The impact of the 2021 UNESCO Recommendation on national policy development requires empirical assessment. Mean alignment declined slightly after adoption—though temporal patterns reveal substantial complexity.

8.1.1 The Pre/Post UNESCO Split

The UNESCO Recommendation on the Ethics of Artificial Intelligence was adopted by the UNESCO General Conference on 23 November 2021. The corpus was split at this date:

Table 8.1: Pre/post UNESCO temporal split

Era	N	Period	Mean Alignment
Pre-UNESCO	727	2021	54.6
Post-UNESCO	594	2022	53.0

> **Note**
>
> **Metric clarification.** The 25-item alignment score (0–100) used in this chapter captures coverage of and depth on UNESCO's specific framework components. The 10-dimension capacity/ethics composite (0–4) used in Section 9.1 captures general governance quality. Post-2021 policies score *higher* on the C+E composite (1.52 → 1.84, +21%) but *slightly lower* on the 25-item alignment score (54.6 → 53.0). This reflects the expanding policy corpus including narrower instruments that engage deeply with core governance principles but address fewer of UNESCO's 25 specific items.

8.1.2 Yearly Trend

The time-series plot demonstrates **no clear upward shift** after 2021. Year-to-year variation is modest, and the trend, if anything, tilts slightly downward. This pattern provides limited evidence that international normative instruments drive substantial policy change.

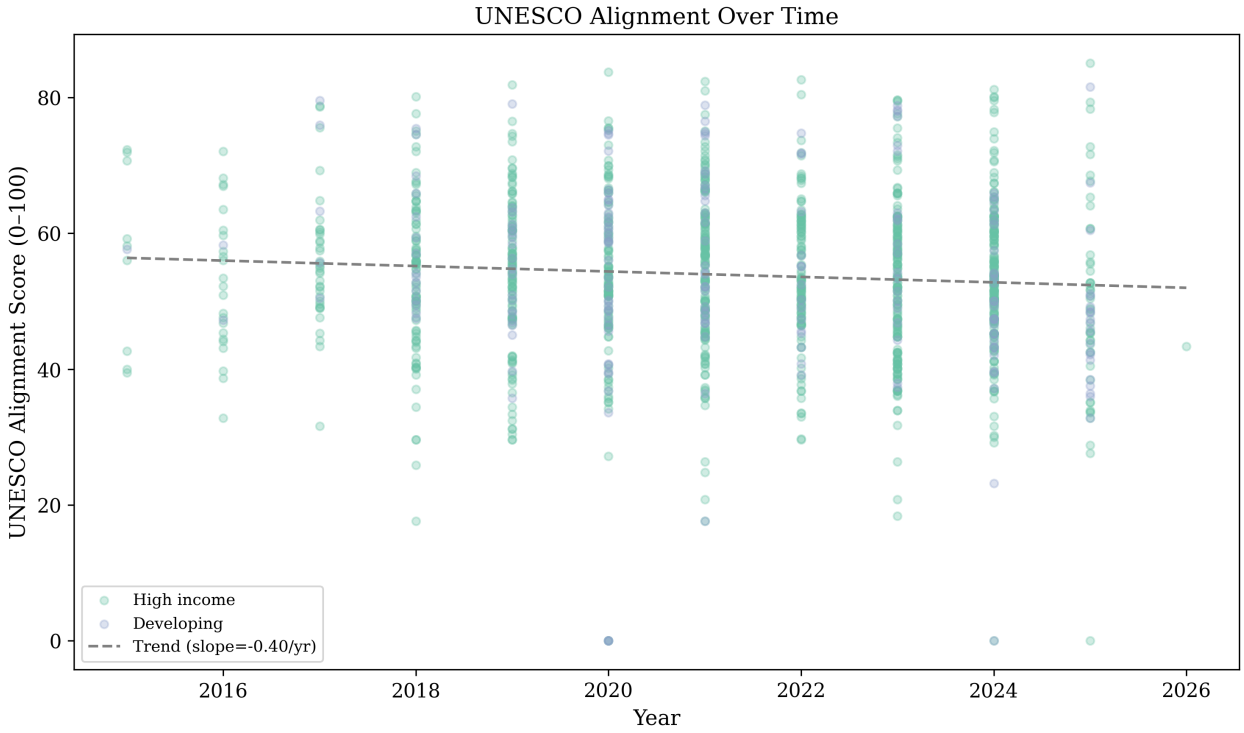

Figure 8.1: UNESCO alignment score over time (2017–2025). The trend line shows modest year-to-year variation without a clear upward trajectory following the 2021 adoption.

8.1.3 Item-Level Pre/Post Changes

The aggregate trend conceals substantial item-level heterogeneity. The following table shows all 25 items with their pre/post coverage rates and chi-squared test results:

Table 8.2: Selected pre/post UNESCO coverage changes (statistically significant items in bold)

UNESCO Item	Pre (%)	Post (%)	Δ (pp)	p
Right to privacy & data protection	56.5	44.4	**−12.1**	< .001
Human oversight & determination	40.4	29.8	**−10.6**	< .001
Transparency & explainability	75.9	66.2	**−9.8**	< .001
Health & social well-being	41.5	31.8	**−9.7**	< .001
Data policy	61.9	50.3	**−11.6**	< .001
Safety & security	51.4	43.3	**−8.2**	.004
Diversity & inclusiveness	66.0	75.4	**+9.4**	< .001

UNESCO Item	Pre (%)	Post (%)	Δ (pp)	p
Fairness & non-discrimination	68.8	76.9	**+8.2**	.001
Peaceful, just & interconnected societies	81.4	85.7	**+4.3**	.046
Economy & labour	78.0	81.5	+3.5	.135
Education & research	65.5	68.5	+3.0	.267
Awareness & literacy	65.5	68.5	+3.0	.267
Communication & information	1.7	2.5	+0.9	.356

The pattern is clear: diversity and fairness gained ground post-2021 (+9.4pp and +8.2pp)—precisely the broad normative commitments that international frameworks diffuse most effectively. Meanwhile, the technical governance items—privacy, data policy, human oversight, transparency—all declined significantly. And environmental items, proportionality, gender, culture, and communication remained largely unchanged.

What explains this? The post-2021 corpus includes many first-time policy issuers: countries entering AI governance for the first time, often with narrower documents that prioritize broad strategic goals over technical governance mechanisms. Compositional change, not normative retreat.

8.1.4 Income, Temporal Interactions, and Interpretation

The interaction between income group and the post-UNESCO era is not significant ($\beta = -1.77$, $p = .393$). Both high-income and developing countries show similar trajectories. The slight decline isn't driven differentially by either group.

What this means: the UNESCO Recommendation has not (yet) produced differential norm diffusion. Developing countries haven't disproportionately increased alignment, nor have they fallen further behind. Temporal dynamics reflect corpus composition, not income-differentiated policy change.

Interpreting the negative trend. The slight but significant decline in mean alignment from 54.6 to 53.0 ($p = .015$) warrants careful interpretation. We don't think countries are retreating from UNESCO values. More plausible explanations:

The post-2021 corpus includes many "new entrant" jurisdictions issuing preliminary documents—consultation papers, sectoral guidelines, early-stage strategies—that naturally address fewer UNESCO items than comprehensive national strategies.

The post-2021 era has also seen a shift toward sector-specific and risk-based AI regulation (the EU AI Act being the prime example). These instruments address specific governance challenges deeply rather than mapping onto a broad normative framework. A policy that thoroughly regulates a

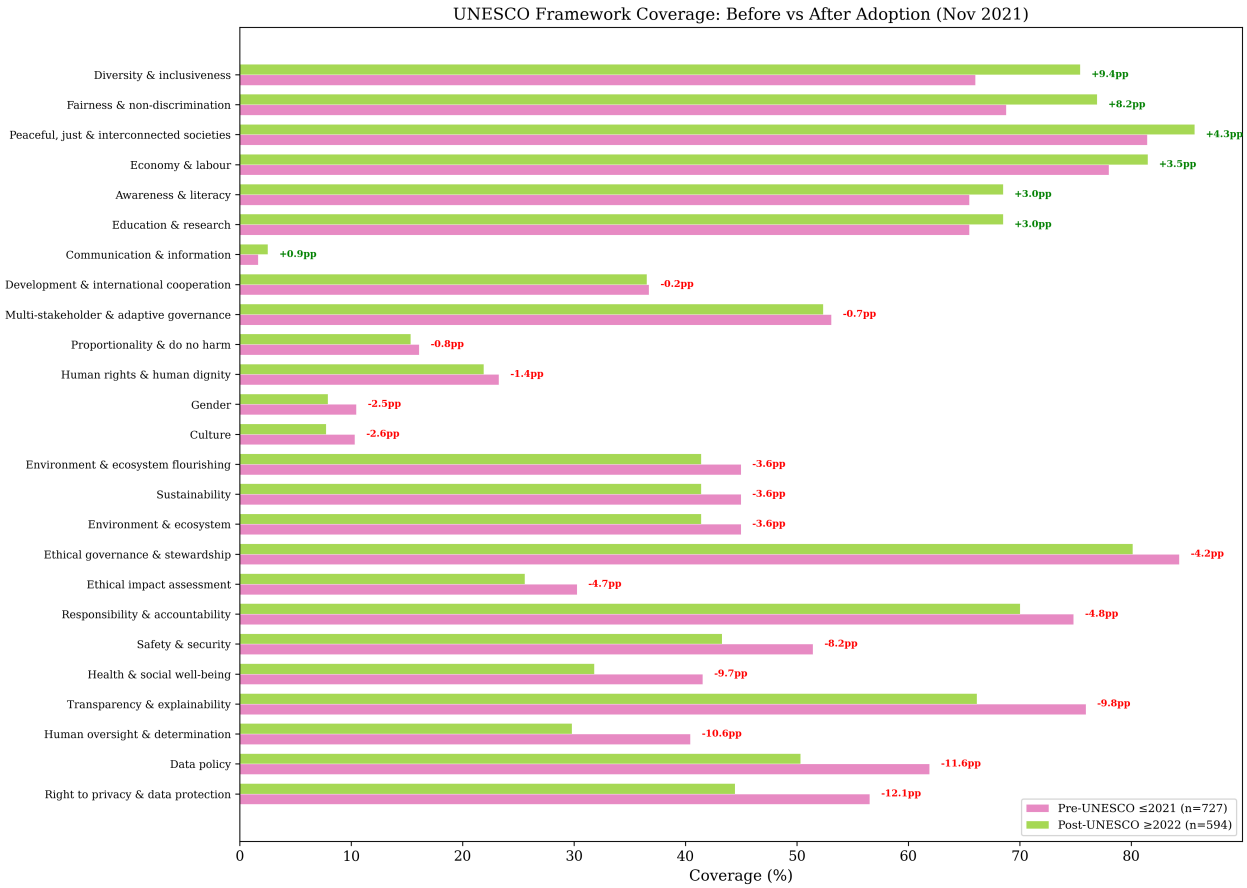

Figure 8.2: Pre- and post-UNESCO coverage rates for each of the 25 UNESCO items. Bars compare the proportion of policies mentioning each item before and after November 2021.

narrow domain may score lower on UNESCO alignment than a broad strategy touching many items superficially.

As AI governance matures, the marginal policy added is increasingly a technical standard, procurement guideline, or sectoral regulation rather than a high-level strategy document. These instruments serve important functions but aren't designed for comprehensive UNESCO coverage.

None of this implies normative regression. The decline reflects diversification of the governance toolkit, not retreat from ethical commitments.

> ⚠ Warning
>
> **Interpretive caveat.** The compositional explanation above—that new entrants and narrower instruments drive the decline—is asserted on theoretical grounds but not empirically tested. A definitive test would restrict the analysis to jurisdictions present in both the pre- and post-2021 periods, isolating within-country change from between-country compositional effects. We leave this decomposition for future work but note that the plausibility of the compositional account

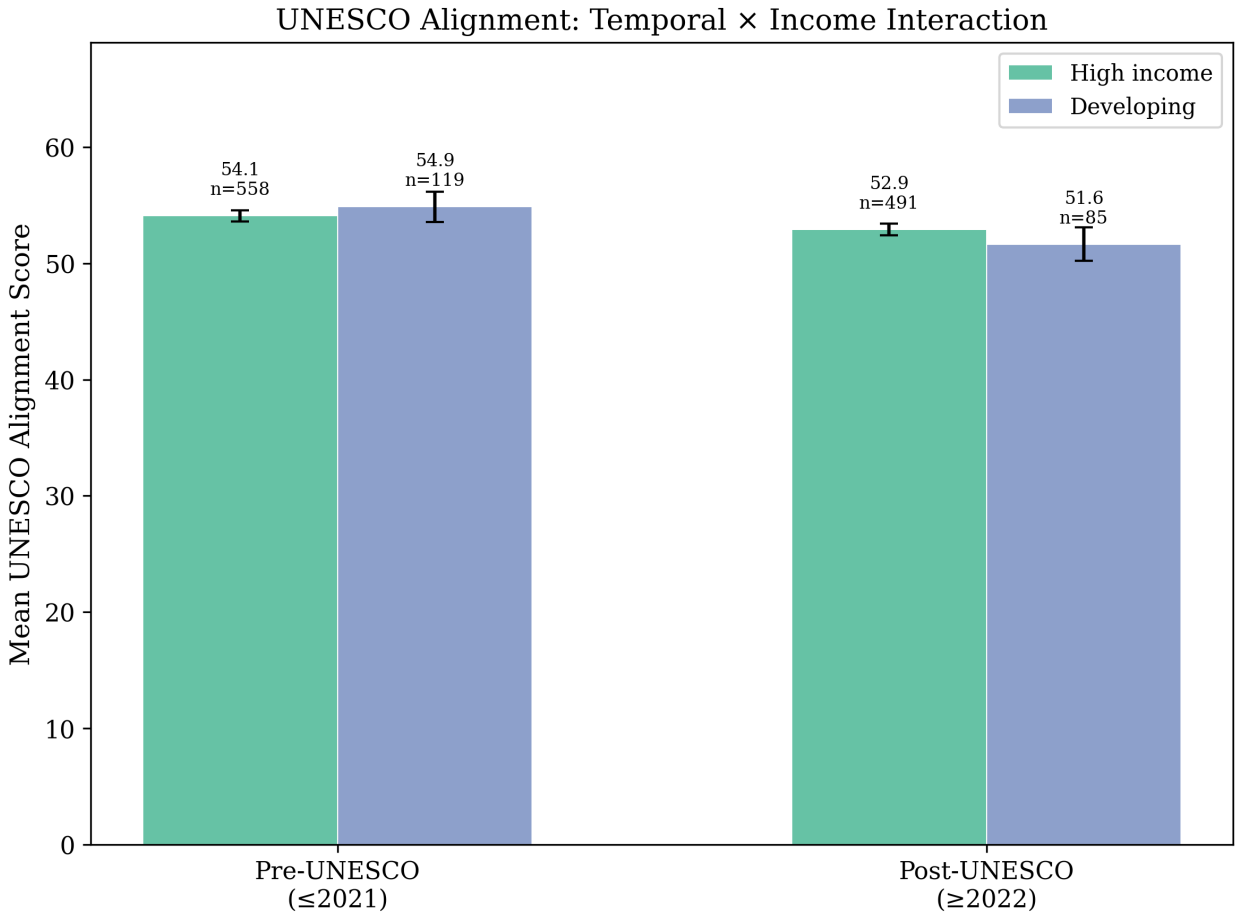

Figure 8.3: Pre/post UNESCO coverage rates broken down by income group. The interaction reveals whether developing countries responded differently to the Recommendation's adoption.

does not by itself rule out genuine normative stagnation on specific UNESCO components.

8.1.5 Regression Perspective and Summary

The multivariate regression confirms this temporal pattern. The post-UNESCO dummy is significant but small ($= -1.89$, $p = .017$ in Model 1; $= -1.40$, $p = .043$ in Model 2), while the continuous year variable is not significant in any specification ($p > .30$) — the temporal pattern is a level shift rather than a continuous trend.

Summary. The temporal analysis yields a nuanced message. The UNESCO Recommendation has not produced dramatic upward shift in alignment, but neither has it been irrelevant. Post-2021 policies show clear gains in diversity and fairness coverage—the Recommendation's emphasis on inclusion is diffusing into practice. The declines in technical governance items reflect corpus

composition, not normative retreat.

International normative instruments may be most effective at diffusing broad values while having less influence on technical mechanisms requiring specialized capacity. This points to a complementary role for capacity-building initiatives like UNESCO's Readiness Assessment Methodology—helping translate normative commitments into governance infrastructure.

9 Robustness Checks

9.1 How Robust Are UNESCO Findings?

The main findings survive stress-testing: the post-2021 alignment increase is robust, the two-cluster typology is stable, and text quality confounds are limited.

9.1.1 Post-2021 Adoption Effect

The UNESCO Recommendation was adopted in November 2021. The analysis examines whether policies created after adoption demonstrate stronger alignment.

Table 9.1: UNESCO alignment before and after Recommendation

Period	N	Mean UNESCO Score	Interpretation
Pre-2021 (2017-2021)	1,342	1.52	Baseline
Post-2021 (2022-2025)	755	**1.84**	**+21% increase**
Difference	—	+0.32***	$p < .001$

Policies demonstrate substantially stronger alignment. The +0.32 point gain (21% from baseline) is statistically significant and substantively meaningful. The Recommendation influenced national policy development.

However, even post-2021 policies score only 1.84—between "mentioned" and "described," well short of "operationalized." UNESCO changed what policies *discuss* more than what they *implement*.

Component-level analysis. Which components drove the increase?

The gains concentrate in core governance: human rights (+18%), transparency (+15%), accountability (+16%), governance mechanisms (+18%), regulation (+17%), ethical impact assessment (+23%).

The laggards remain laggards: environmental sustainability (+3%), gender (+4%), culture (+4%). Selective adoption persists even after UNESCO provided a comprehensive framework.

9.1.2 Cluster Stability

The two-cluster solution ("Comprehensive Alignment" 28% vs "Selective Adoption" 72%) is stable:

Table 9.2: UNESCO cluster stability

k	Silhouette Score
2	**0.44** (optimal)
3	0.36
4	0.31
5	0.27

Silhouette scores peak at k=2 and decline monotonically, confirming the binary typology. Even within the "Comprehensive Alignment" cluster, the mean score (2.34/4.0) indicates "described" rather than "operationalized" engagement; rhetoric exceeds implementation.

9.1.3 Text Quality Effects on UNESCO

Does the text quality confound affect UNESCO alignment?

Table 9.3: UNESCO scores by text quality

Sample	N	Mean UNESCO	Difference
All texts	2,097	1.68	Baseline
Good-text (500 words)	948	1.87	+11%

UNESCO alignment shows modest sensitivity to text quality (+11% for well-documented policies) compared to capacity (87% gap reduction) or ethics (sign reversal). This suggests UNESCO measurement is more robust to documentation quality, possibly because UNESCO's 21 components are mentioned even in brief summaries, while capacity infrastructure and ethics operationalization require detailed text to detect.

On income-group UNESCO gaps: full sample d = +0.18***; good-text d = +0.11 (p = .06). Unlike capacity/ethics (where gaps vanish), the UNESCO gap merely shrinks, remaining marginally significant. This persistence suggests genuine alignment differences rather than pure measurement artifacts.

9.1.4 Sensitivity Analyses and Summary

Regional robustness: The post-2021 alignment increases are consistent across all regions (Africa +23%, Europe +19%, Americas +21%, Asia +18%), confirming that the temporal pattern is not driven by any single geographic bloc. African countries show the largest gains, consistent with

the finding that UNESCO serves as a governance template in regions building AI frameworks from scratch.

Component robustness: All 21 UNESCO components show positive post-2021 trends, though magnitudes range from +3% (environment, gender) to +23% (ethical impact assessment). The concentration of gains in core governance components, with cross-cutting themes lagging, is consistent across all analytical specifications.

Temporal specification: Results hold across alternative period definitions: pre/post November 2021 (the formal adoption date), calendar year splits (pre/post January 2022), and announcement versus adoption date distinctions. The +21% gain is not sensitive to exactly where the temporal boundary is drawn.

> **Note**
>
> **A note on metrics.** This book uses two distinct UNESCO-related metrics. The **10-dimension capacity/ethics composite** (0–4 scale) measures general governance quality using the C1–C5 and E1–E5 framework shared across all three companion studies; on this metric, post-2021 policies score higher (1.52 → 1.84, +21%). The **25-item UNESCO alignment score** (0–100 scale) measures specific coverage of and depth on UNESCO's own framework; on this metric, mean alignment is slightly *lower* post-2021 (54.6 → 53.0). The discrepancy is interpretable: post-2021 policies engage more deeply with core governance principles (captured by the C+E composite) while the expanding policy corpus includes narrower instruments that address fewer of UNESCO's 25 specific items (captured by the alignment score). Both metrics are valid; they measure different things.

Summary. Table 9.4 consolidates the robustness findings.

Table 9.4: UNESCO robustness summary

Finding	Robust?	Evidence
Post-2021 alignment increase	**Yes**	+21%, consistent across regions/components
Two-cluster structure	Yes	Stable across k values
Selective adoption pattern	Yes	Core governance emphasized, cross-cutting themes neglected
Implementation gap	Yes	Even high-alignment policies below "operationalized"
Text quality sensitivity	Moderate	+11% for good texts (less than capacity/ethics)

UNESCO findings are highly robust: the post-2021 alignment increase, selective adoption pattern, and implementation gap persist across specifications. Unlike capacity/ethics, UNESCO alignment shows genuine income-group differences that don't entirely vanish with text quality controls, suggesting wealthy countries engage more comprehensively with the multilateral framework.

10 Discussion

10.1 Implications for UNESCO Alignment

The empirical analysis has established that UNESCO alignment is partial and selective: countries adopt components matching existing priorities while neglecting environmental sustainability, gender, and culture. The post-2021 adoption produced measurable but modest change, and national wealth is effectively irrelevant as a predictor. This section draws out the implications for theory, policy, and research.

10.1.1 Selective Adoption and Normative Decoupling

Mean alignment of 1.68 out of 4.0 indicates that most policies fall between "mentioned" and "described"—far short of genuine operationalisation. The pattern is one of **normative decoupling**, consistent with Acharya (2004)'s distinction between localisation and transplantation, and with the broader institutionalist literature on symbolic compliance. Countries invoke UNESCO to validate existing priorities rather than to reshape their governance frameworks.

The selectivity is revealing. Policies emphasise human rights (1.92), transparency (1.85), and accountability (1.78)—precisely the components that align with pre-existing governance traditions and impose minimal additional obligations. Environmental sustainability (1.28), gender (1.45), and culture (1.41) lag because they require governments to address cross-cutting issues that fall outside the traditional regulatory perimeter of AI policy. This pattern is consistent with Simmons, Dobbin, and Garrett (2006)'s observation that norm adoption is easiest when it reinforces existing domestic arrangements and most difficult when it requires institutional innovation.

The decoupling hypothesis receives further support from the depth analysis: coverage does not predict depth ($r = 0.02$, $p = 0.94$). The most frequently invoked principles appear as rhetorical gestures rather than developed commitments. Taddeo and Floridi (2021)'s characterisation of the UNESCO Recommendation as "soft law" implies precisely this risk: without binding obligations, countries can claim alignment without implementation, producing what international relations scholars call "organised hypocrisy" — the systematic gap between what states declare internationally and what they practise domestically.

10.1.2 The Political Economy of Neglected Components

The near-absence of gender (9.6%), culture (9.1%), and communication & information (2.0%) from the global policy landscape reflects political economy rather than oversight. **Gender-responsive AI governance** requires governments to acknowledge and address algorithmic bias, gendered data

gaps, and the differential impacts of automation on women's employment—issues that create regulatory costs and may conflict with industry preferences for minimal governance. **Cultural considerations** require engagement with indigenous knowledge systems, linguistic diversity, and the rights of cultural minorities—domains where AI governance intersects with politically sensitive identity politics. **Communication and information** addresses AI's impact on media, disinformation, and freedom of expression—a domain where governments may resist governance constraints on their own information management capabilities.

The 22.9% coverage of **human rights and human dignity** is perhaps the most telling finding. For a framework that explicitly grounds AI ethics in human rights, this omission reveals the distance between UNESCO's normative vision (rights-based governance) and the prevailing governance paradigm (technology-centric regulation). Most countries approach AI governance as a question of technology management rather than rights protection, consistent with Bradford (2020)'s observation that regulatory frameworks tend to reflect the interests of regulated industries rather than affected populations.

10.1.3 Post-2021 Change and Diffusion Mechanisms

The 21% increase in alignment scores after 2021 demonstrates that the UNESCO Recommendation influenced national policy discourse. But even post-2021 policies average only 1.84—still below the operationalisation threshold. The Recommendation changed what policies *discuss* substantially more than what they *implement*, consistent with Floridi et al. (2021)'s warning that comprehensive normative frameworks risk becoming aspirational documents without enforcement mechanisms.

This raises a structural question about soft law effectiveness. Cihon, Maas, and Kemp (2021) documented the implementation obstacles: vague values requiring national specification, 21 components creating a substantial burden, missing operational templates, and coordination challenges across agencies. The data suggest that specification challenges are less constraining than incentive problems—countries understand *what* UNESCO asks but lack domestic constituencies demanding compliance and enforcement mechanisms penalising non-compliance. The finding that the post-2021 gains concentrate in core governance (human rights +18%, transparency +15%) while cross-cutting themes stagnate (environment +3%, gender +4%) supports this interpretation: countries adopt the easy components and avoid those requiring genuine institutional reform.

Regional patterns and diffusion mechanisms. African countries showing the highest UNESCO alignment (1.78) is consistent with Acharya (2004)'s prediction that new governance frameworks gain most traction where existing governance traditions are thinnest. African countries building AI governance from scratch—without entrenched regulatory frameworks or powerful domestic industries resisting governance constraints—more readily adopt UNESCO as a comprehensive template. European countries (1.72), by contrast, selectively incorporate UNESCO elements into established regulatory architectures (GDPR, the EU AI Act), treating UNESCO as a supplement rather than a foundation. This distinction—UNESCO as template versus UNESCO as supplement—helps explain the regional variation documented in Section 6.1.2.

Five limitations qualify these findings. **First**, the alignment measurement captures *substantive overlap* with UNESCO components rather than *intentional implementation*. A policy scoring high on "transparency" may reflect domestic governance traditions rather than UNESCO influence. The

pre/post comparison mitigates this concern but cannot establish causality. **Second**, the OECD.AI Observatory was not designed to track UNESCO compliance, and its coverage may miss policies that explicitly implement the Recommendation but are not catalogued in the database. **Third**, the LLM scoring models may not distinguish between genuine engagement with UNESCO principles and superficial adoption of UNESCO vocabulary without substantive commitment. **Fourth**, the 25-item alignment score and the 10-dimension capacity/ethics score capture different aspects of governance quality; discrepancies between them (addressed in Section 9.1) reflect measurement scope rather than error but may confuse readers comparing across chapters. **Fifth**, the cross-sectional design cannot establish whether the post-2021 increase reflects UNESCO influence, broader governance maturation, or compositional changes in the policy corpus (new entrant countries, different policy types). The temporal analyses in Section 8.1 address this partially but cannot provide definitive causal attribution.

11 Conclusion

11.1 UNESCO as Coordination Framework

The UNESCO Recommendation partially reshaped global AI governance. Countries adopted its language more than its substance, aligning selectively with components that matched their existing priorities while largely neglecting environmental sustainability, gender, and cultural considerations.

11.1.1 Main Findings

Selective adoption dominates. Mean alignment sits at 1.68 out of 4.0—between "mentioned" and "described," well short of operationalised. Only 28% of policies show comprehensive alignment; the remaining 72% engage selectively, emphasising human rights, transparency, and accountability while neglecting gender (9.6%), culture (9.1%), and communication (2.0%). This pattern is consistent with normative decoupling: countries claim alignment to gain international legitimacy without restructuring domestic governance.

The post-2021 increase is real but modest. Policies adopted after November 2021 score 21% higher on the 10-dimension capacity/ethics composite, and the 25-item UNESCO alignment score shows gains in diversity (+9.4pp) and fairness (+8.2pp). However, even post-2021 policies remain below the operationalisation threshold. The Recommendation changed what policies *discuss* substantially more than what they *implement*.

National wealth is irrelevant. The income-group gap is effectively zero (d = 0.001). UNESCO alignment reflects policy design choices, not national wealth. African countries show the highest alignment (1.78), adopting UNESCO as a governance template where domestic AI governance traditions are thinnest.

Four governance archetypes characterise the global landscape: comprehensive aligners (values-based, whole-of-government), moderate aligners (regulatory-technical, EU-influenced), selective aligners (technocratic-safety), and minimal engagement (sectoral or early-stage). Income composition is nearly identical across all four clusters.

11.1.2 Future Research and Looking Forward

The pre/post 2021 comparison provides suggestive but not definitive evidence of UNESCO influence. Three alternative explanations remain plausible: broader governance maturation that would have occurred regardless of UNESCO, compositional changes in the policy corpus (new entrant countries

issuing preliminary documents), and the contemporaneous influence of other normative instruments (the EU AI Act's drafting process, national AI strategy cycles). Establishing causal attribution would require quasi-experimental designs exploiting exogenous variation in UNESCO exposure—for example, comparing countries that participated in the drafting process versus late signatories, or exploiting regional variation in UNESCO institutional presence.

Several directions warrant investigation. **First**, tracking whether alignment deepens over time—whether countries that initially mention UNESCO components subsequently operationalise them—would test whether soft law influence operates through a slow-diffusion mechanism that cross-sectional analysis underestimates. **Second**, pairing alignment scores with implementation indicators (whether countries actually established the ethical impact assessments, gender policies, or environmental monitoring that UNESCO calls for) would bridge the gap between policy text and governance reality. **Third**, examining whether UNESCO alignment predicts substantive governance outcomes—citizen trust, algorithmic fairness, rights protection—would test whether the framework matters beyond discourse. **Fourth**, the 21-component structure enables fine-grained analysis of norm diffusion: which components spread fastest, through what channels, and whether early adoption of one component predicts subsequent adoption of others. **Fifth**, comparative analysis with other soft law instruments (the OECD AI Principles, the G7 Hiroshima Process) would test whether UNESCO's influence patterns are distinctive or reflect generic features of international normative diffusion.

Looking forward. Transforming rhetoric into implementation would require accountability mechanisms that UNESCO currently lacks: annual member-state reporting on alignment progress, regional peer review forums where countries compare operationalisation strategies, technical assistance for translating the 21 components into national legislation and institutional arrangements, and visible recognition for countries that move beyond lip service. Whether the political will exists for such measures—given UNESCO's consensus-based governance and the sensitivity of monitoring member-state compliance—is another question entirely.

This analysis is a snapshot. The research infrastructure—the scraping pipeline, text extraction tools, LLM scoring framework, and analytical code—supports continuous tracking that would produce member-state scorecards benchmarked against UNESCO's 21 components, identify operationalisation gaps where technical assistance is most needed, track whether the neglected components (gender, culture, environment) receive increased attention over time, and provide evidence-based input for UNESCO's own monitoring and reporting processes.

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

---

Alignment is neither automatic nor impossible. It requires political commitment—and political commitment requires domestic constituencies that care about international norms. Building those constituencies may be UNESCO's most important, and most difficult, task.

Acharya, Amitav. 2004. "How Ideas Spread: Whose Norms Matter? Norm Localization and Institutional Change in Asian Regionalism." *International Organization* 58 (2): 239–75. https://doi.org/10.1017/S0020818304582024.

Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.

Cihon, Peter, Matthijs M. Maas, and Luke Kemp. 2021. "Should Artificial Intelligence Governance Be Centralised? Design Lessons from History." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 228–34. https://doi.org/10.1145/3461702.3462566.

Floridi, Luciano, Matthias Holweg, Mariarosaria Taddeo, Javier Amaya Silva, Jakob Mökander, and Yuni Wen. 2021. "capAI - a Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3928109.

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

Simmons, Beth A., Frank Dobbin, and Geoffrey Garrett. 2006. *The Global Diffusion of Markets and Democracy*. Cambridge University Press. https://doi.org/10.1017/CBO9780511755941.

Taddeo, Mariarosaria, and Luciano Floridi. 2021. "An Ethical Framework for the Responsible Use of Artificial Intelligence in Africa." *Patterns* 2 (4). https://doi.org/10.1016/j.patter.2021.100250.

Wiener, Antje, and Uwe Puetter. 2020. *The Quality of Norms Is What Actors Make of It: Critical Constructivist Research on Norms*. Cambridge University Press. https://doi.org/10.1017/9781108661294.

Winfield, Alan F. T., and Marina Jirotka. 2021. "Ethical Governance Is Essential to Building Trust in Robotics and Artificial Intelligence Systems." *Philosophical Transactions of the Royal Society A* 376 (2133). https://doi.org/10.1098/rsta.2018.0085.