

The Politics of AI Ethics: How Normative Principles Translate into Law and Policy in South Korea

Chee Hae Chung¹, Daniel S. Schiff²

¹ Harvard Kennedy School

² Purdue University

cheehae_chung@hks.harvard.edu, dschiff@purdue.edu

Abstract

A growing global consensus around AI ethics principles has not produced equivalent convergence in law and policy. How normative principles are translated into governance remains understudied: we know little about why some principles survive or disappear as they move into policy, how institutional pathways shape that process, or how translation itself should be measured. We develop a framework for ethics-to-governance translation and identify possible sources of variation: principles' demands, institutional fit and state capacity, policy instrument 'pathways,' and political administration. We examine South Korea, an early mover in comprehensive AI regulation, analyzing 15 ethical principles across 69 government documents issued between 2017 and 2025: 23 ethics guidelines, 27 non-binding executive policy documents, and 19 enacted statutes and enforcement decrees. We measure principle-level keyword prevalence, validated with keyword-in-context and embedding checks, compute a translation gap for binding and non-binding governance pathways, and code whether provisions specify obligations, duty-bearers, procedures, penalties, and remedies. We find that internationally recognized and Korean-salient principles begin at similar levels of prominence in ethics guidelines but diverge substantially as they move into governance, especially binding law. The pattern is not simply related to international versus domestic values: principles such as governance and auditing, which map onto established functions of administrative coordination, monitoring, and verification, survive more than principles requiring less-developed rights enforcement capacities. Binding law also filters normative content more strongly than non-binding policy. Yet when rights-oriented principles such as justice/fairness and transparency reach statutory text, they are comparably operative; the main difference is legal coverage, not drafting strength. These findings are consistent with institutional fit as one explanation for selective translation while leaving room for political, cultural, strategic, and other institutional explanations. More broadly, we conceptualize ethics-to-governance translation as a multidimensional process requiring explanation in its own right and offer a replicable framework for studying which ethical principles shape governance, through which pathways, in what form, and under what conditions.

Introduction

Since 2016, hundreds of organizations have issued ethics guidelines invoking the same principles, among them transparency, fairness, accountability, dignity, and non-maleficence (Corrêa et al. 2023; Jobin, Ienca, and Vayena 2019), yet legal institutionalization has not kept pace and implementation failure persists (Mittelstadt 2019; Hagendorff 2020). Few jurisdictions have enacted comprehensive AI laws, and existing laws often provide vague aspirational language rather than concrete compliance requirements (Kaminski 2023; Veale and Borgesius 2021). The EU AI Act, adopted in 2024, exemplifies this difficulty: even with detailed technical specifications, enforcement arrangements remain uncertain and member state transposition has proven slow (Veale and Borgesius 2021) and delayed by competing industry and international pressures.

The gap is not mere policy lag. Much scholarship treats implementation as a sequence in which principles are articulated, rendered into legal language, enacted, and enforced. Policy process scholars have long questioned this model as descriptively inadequate, since causal mechanisms are under-specified and the stages are rarely linear in practice (Sabatier 2007). We suggest that translation of ethics into policy and practice depends on many factors, including problem definition, political ideology, industry and organizational interests, technical feasibility, and framing, amongst others (Baumgartner et al. 2009; Entman 1993; Kingdon 2011; Mittelstadt 2019; Sabatier 2007; Stone 1989).

Examining all of these is beyond the scope of one paper; we take up one piece, arguing that translation should depend at least in part on the compatibility between what principles demand of the state and what the state's existing institutional apparatus is configured to deliver. For instance, principles

requiring state-led management and oversight might translate readily where such capacity exists, whereas principles requiring the state to create novel rights-enforcement mechanisms for individuals or new collective socio-technical imaginaries might meet more resistance (Carpenter 2002; Jasanooff 1998), reflecting differences in technical feasibility and political acceptability known in the agenda-setting literature (Kingdon 2011). We thus treat institutional fit as one important explanation rather than a determining one.

South Korea provides a strategically valuable case for testing this argument. Korea is one of the first regulatory movers in AI and has an institutionally layered history: its developmental trajectory left a highly capable, state-led bureaucracy accustomed to steering technological change and equipped with strong coordination mechanisms (Amsden 1989; Evans 1995; Wade 2018; Woo-Cumings 2019), and it enacted one of the first comprehensive national AI statutes outside the EU, while democratic consolidation since 1987 encouraged greater stakeholder consultation and legislative deliberation (Im 2004). Meanwhile, its administrative traditions likewise emphasize principles like collective welfare and hierarchical accountability (Bell 2010; Wong 2020).

Where administrative capacity is so prioritized, translation failures are more plausibly consequences of structural design rather than a symptom of overall weak institutions. This makes Korea a strong case for studying the importance of institutional fit (Eckstein 2000; Gerring 2006). If translation gaps track what the bureaucracy is organized (or not organized) to do where capacity is this high, the same logic should hold elsewhere; for instance in high-capacity developmental states such as Taiwan and Singapore.

A small empirical literature documents ethics-to-law gaps directly. Jobin, Ienca, and Vayena (2019) and Corrêa et al. (2023) show that guidelines converge on a few principles while rarely specifying enforcement, and literature acknowledges that the wording of principles is itself strategic, shaping political reception and legal force (Entman 1993; Stone 1989). Djeflal, Siewert, and Wurster (2022) similarly find frequent invocation but minimal regulatory specification across national AI strategies. Closest to ours, Lee and Lee (2025) compare Korean ethics guidelines with legislative bills and find transparency and accountability prominent in guidelines but thinly covered in bills. These studies establish that gaps exist but do not explain why particular principles fare better than others.

Many questions therefore remain open: What does it mean for a principle to be ‘translated’ into policy, and how should we weigh subtle changes in language that may change the meaning or force of principles? When does manifestation in text-as-discourse translate to eventual policy action, and how can ‘meaningful implementation’ be empirically measured? Does a principle’s fate depend on its abstract normative appeal and ability to achieve consensus, its fit with a country’s extant values and what the state is already administratively organized to do, or still other factors?

We approach these through three guiding questions. First, does translation depend on a principle’s origin,

whether internationally recognized (Fjeld et al. 2020) or salient in Korean governance? Second, does it depend on the instrument, so that principles translate differently into binding law or executive action than into non-binding policy? Third, does translation differ across political ideologies and administrations? We address these through quantitative and qualitative content analysis of 69 South Korean AI governance documents spanning ethics guidelines (N = 23), policy frameworks (N = 27), and enacted laws and presidential decrees (N = 19), issued between 2017 and 2025 across both the liberal Moon Jae-in (2017-2022) and conservative Yoon Suk-yeol (2022-2025) administrations.

While this account cannot cover all the dimensions of translation and its measurement, or of the many possible predictors of translation across diverse cultural and governmental settings, it offers a set of techniques as an initial framework for studying ethics-to-policy translation.

Literature Review

The Global AI Ethics Consensus

AI ethics guidelines have proliferated since 2016. An analysis of 84 guidelines identified five recurring principles, transparency, justice/fairness, non-maleficence, responsibility, and privacy (Jobin, Ienca, and Vayena 2019); a broader mapping extended these to 11 (Fjeld et al. 2020); a review of 200 guidelines through 2023 confirmed the convergence (Corrêa et al. 2023); and a review across public, private, and NGO sectors studied 25 distinct principles, finding shared priorities but variation across sectors (Schiff et al. 2021).

This convergence reflects isomorphic pressures within the transnational AI ethics community (DiMaggio and Powell 1983; Haas 1992; Hagendorff 2020). However, guidelines vary dramatically in specificity, and many lack enforcement mechanisms (Floridi 2018; Mittelstadt 2019; Zhu and Xu 2023) or even explicit regulatory engagement (Schiff et al. 2021), and principles often conflict in ways guidelines rarely address (Whittlestone et al. 2019).

Moreover, scholars have questioned what this consensus represents. Because the guidelines are overwhelmingly authored by well-resourced actors in the Global North, their convergence may reflect agenda-setting power rather than fully considered or participatory agreement on values (Jobin, Ienca, and Vayena 2019; Png 2022). And because it is an agreement over vocabulary rather than meaning, with contested terms like fairness and transparency carrying different implications across institutional and cultural settings (Mulligan et al. 2019; Whittlestone et al. 2019), we might expect principles to diverge as they enter national law. This paper therefore also implicates a rich scholarship on the question of universal, international, or ‘harmonized’ ethics against contextually, linguistically, or culturally-specific ethics (Mohamed, Png, and Isaac 2020; ÓhÉigeartaigh et al. 2020).

The Ethics-to-Law Translation Gap

Legal institutionalization remains far more limited. The EU AI Act (2024), the most comprehensive effort to date, faces significant implementation uncertainties (Veale and Borgesius 2021), and while national AI strategies invoke ethical principles widely, there is far less confidence about how seriously those concepts bind. Identified impediments include the absence of compliance requirements and accountability mechanisms (Djeffal, Siewert, and Wurster 2022), temporal misalignment between technological change and legislation (Larsson 2020), institutional fragmentation (Calo 2017), and lack of substantive carry through between policy ‘mission statements’ and actual implementation details (Schiff 2023).

Recent empirical studies document translation gaps more systematically, though most prior work covers a single sector, country, or document type. Closest to this paper, Lee and Lee (2025) use large language models to compare Korean ethics guidelines with legislative bills and find transparency and accountability prominent in guidelines but thinly covered in bills. Ghimire and Edwards (2024) report the same disconnect in education policy, and Larsson (2020) characterizes European ethics guidelines as largely symbolic, letting governments appear responsive without costly regulatory action. These studies establish that translation failures are real and widespread.

A parallel literature questions whether national AI documents primarily serve to advance ethical vision at all, reading them instead as signaling efforts aimed at investors, international allies, and rivals (Bareis and Katzenbach 2022). A principle’s presence may therefore index economic or political positioning rather than commitment, a possibility captured by typologies that distinguish signaling motivations, such as signaling social responsibility or signaling leadership, from substantive goals (Schiff et al. 2022).

To assist with this question of signaling versus substance, we therefore separate documents by legal force rather than by branch of government, and articulate these as two distinct pathways. Both branches of government in Korea we study, legislative and executive, issue binding and non-binding policy: enacted statutes and presidential enforcement decrees create obligations that courts can enforce, whereas strategic plans, roadmaps, and implementation guidelines direct government activity without binding force. Our ‘law translation pathway’ therefore groups statutes with enforcement decrees, while the ‘non-binding policy pathway’ covers only these non-binding instruments.

Theoretical Gaps and Analytical Framework

In sum, existing scholarship documents aggregate ethics-to-law gaps but rarely asks whether some principles translate more readily than others, leaving cross-principle variation largely unexamined. However, preliminary evidence sug-

gests heterogeneity: privacy protection has achieved substantial regulatory institutionalization across jurisdictions while protections rooted in dignity and solidarity, principles prominent in international and European declarations but without an established administrative home in Korean practice, arguably remain largely aspirational (Hildebrandt 2020; Kaminski 2023).

Criticism of AI principlism is extensive (Mittelstadt 2019; Whittlestone et al. 2019), but it typically targets the guidelines themselves rather than regulation, and has not engaged with what states are actually equipped to do. Understanding the sources of this variation, whether stakeholder power, technical feasibility, or cultural fit, therefore matters both for explaining key patterns in translation and for designing governance or institutional reform that can realize the aims of ethical principles and guidelines.

The factors we foreground here, amongst many others, are institutional fit and capacity—what a government is already organized to do. We argue that technology governance scholarship suggests that institutional capacity should matter more than normative sophistication for real-world outcomes: biotechnology governance, GDPR implementation, and algorithmic impact assessments succeeded only where pre-existing governmental arrangements provided a foundation (Jasanoff 2003; Kaminski and Malgieri 2021; Reisman et al. 2018). We therefore propose that translation success depends (at least in part) on the compatibility between what principles demand of the state and what the state’s institutional apparatus can deliver.

This explanation is institutional as well as cultural. Principles requiring state-led management, oversight, and verification, functions developmental states like Korea are configured to perform should translate readily, while principles requiring new individual rights-enforcement machinery should meet greater resistance. Institutional fit is not the only candidate explanation, however: outcomes also reflect interest-group power (Baumgartner et al. 2009), problem framing (Entman 1993; Stone 1989), technical feasibility (Kingdon 2011), and a deliberate preference for vague commitments that preserve discretion (Sunstein 1995). We return to these briefly in the discussion.

A final empirical and conceptual challenge is that translation admits of degrees. A principle may be present or absent; merely declaratory or substantive; named with or without attaching to any responsible actor; and containing or lacking a procedure, penalty, remedy. The design of specific policy tools, instruments, and mixes is therefore consequential in assessing translation, and likely requires significant research in specific settings, sectors, and jurisdictions.

We therefore extend the principles-to-policy literature by distinguishing two governance pathways or ‘outcomes,’ both of them represented in law-or-policy documents rather than ethics frameworks. The binding pathway (enacted stat-

utes and enforcement decrees) carries legal force enforceable through the courts; the non-binding pathway (strategic plans, roadmaps, and implementation guidelines) directs administrative activity without creating enforceable obligations. Because Korean enforcement decrees are delegated legislation that binds, bindingness cuts across the legislative-executive divide rather than tracking it.

We expect the binding pathway to produce larger translation gaps, because drafting conventions, parliamentary negotiation, constitutional review, and the need to render abstract norms enforceable should filter normative content more aggressively than aspirational frameworks do. We further expect Korean-salient principles, such as governance, auditing, public good, and harmony, which name state-led management and oversight functions aligned with Korea's developmental legacy, to translate more effectively than internationally recognized principles, which require rights-enforcement machinery with less precedent. Finally, we assess translation across two different governmental administrations but present this as a more exploratory analysis.

Principles differ in what they ask the state to do, and we treat that difference as the operative mechanism. Rights-oriented principles ask the state to create, adjudicate, or enforce entitlements held by individuals against public or private actors: these include justice/fairness, transparency, privacy, dignity, and freedom/autonomy. Management-oriented principles instead ask the state to organize, monitor, and verify administrative activity: these include governance, auditing, and non-maleficence in its safety-assurance form. The remaining principles we examine, including public good, harmony, beneficence, trust, sustainability, solidarity, and responsibility, name diffuse collective goods that require neither an enforcement apparatus nor a specific administrative routine. We assign each of the fifteen principles to one of these three types and expect management-oriented principles to translate most readily, rights-oriented principles least readily, and diffuse principles to pattern with the rights-oriented set because neither has an existing administrative home. Since the internationally recognized and Korean-salient sets differ sharply on this dimension, the origin split serves as an observable proxy for demand type throughout the analysis.

This study makes several theoretical and methodological contributions. We (1) contrast principles prominent in the international consensus with those long salient in Korean governance; (2) examine four Korean-salient principles, governance, auditing, public good, and social harmony, and show that those naming administrative routines translate differently from those naming diffuse collective goods; (3) trace which principles survive into binding law versus non-binding policy; (4) measure ethics-to-policy translation explicitly as a four-level ladder, from absent to declaratory to operationalized to enforceable; and (5) provide an initial replicable measurement framework that other countries, administrations, and issues can reuse to study translation.

Research Design and Methodology

Korean Value Context

We label four principles as “Korean-salient:” governance, auditing, public good, and harmony. They are not unique to Korea and overlap in various ways with internationally-recognized principles; what sets them apart is their prominence in, and fit with, Korea's cultural and administrative tradition, where oversight and stewardship of the collective interest are long-standing public values rather than mere technique (Jorgensen and Bozeman 2007): a reader outside the region may not recognize them as ethical principles at all, which is itself informative about how far the international consensus travels. Yet we do not wish to overstate these values as cultural essences, and we note that shared vocabulary can carry different meanings across settings (ÓhÉigeartaigh et al. 2020).

Nevertheless, oversight and accountability serve as Korean-salient ethical principles because they make normative claims about how public power should be exercised and held to account. Local administrative traditions that value collective welfare and hierarchical accountability are also consistent with, and reinforce, this orientation (Bell 2006; Pye and Pye 1985). Institutionally, Korea's administrative law tradition emphasizes *ex-post* monitoring and compliance verification through systematic audit and inspection procedures (Ginsburg 2003; Im et al. 2014). Operationally, the Government Performance Evaluation Act (2001) establishes comprehensive audit frameworks for governmental activities; specialized inspection agencies conduct regular compliance assessments across policy domains.

Document Corpus

Two branches of government produce the documents we study. The National Assembly, Korea's elected legislature, enacts statutes; the executive, comprising the president and the ministries, issues the enforcement decrees that give statutes operative detail and the strategic plans and guidelines that steer administrative activity. Both are understood to share responsibility for guiding technological development, a legacy of Korea's developmental state (Amsden 1989; Woo-Cumings 2019). Specifically, we assembled a corpus of 69 AI governance documents issued between 2017 and 2025 through systematic searches of the Korean Government Legislation Center (law.go.kr), ministry websites (MSIT, MOIS, PIPC, FSC, KCC, MOE), and semi-governmental agency publications (NIA, KISA).

We organize the corpus explicitly around what our analysis compares: where principles start and where they end up. Ethics guidelines (N = 23; principle frameworks, self-assessment checklists, and sectoral ethics standards) serve as the baseline. We then trace principles into two target path-

ways distinguished by legal force: non-binding policy documents (N = 27; strategic plans, roadmaps, implementation plans, and regulatory guidance), which direct administrative activity without creating enforceable obligations, and binding law or executive action (N = 19; enacted statutes, N = 11, and enforcement decrees, N = 8), which does. We include only enacted legislation and currently active policy documents, excluding pending bills, local ordinances, and private sector guidelines, and for amended laws include only the initial enactment and the most recent version. This is the near-complete universe of publicly available national-level AI governance documents for the period.

Principle Framework Development

We analyze 15 ethical principles: 11 internationally recognized principles derived from comprehensive global mapping (Fjeld et al. 2020), encompassing transparency, justice/fairness, non-maleficence, responsibility, privacy, dignity, beneficence, freedom/autonomy, trust, sustainability, and solidarity; plus 4 “Korean-salient” principles identified through preliminary document analysis, Korean law and governance literature (Ginsburg 2003; Im et al. 2014), and comparative East Asian governance scholarship (Bell 2010; Wong 2020). The (translated) Korean-salient principles are: **Auditing/Verification** (*geomjeung/gamsa*): Emphasis on *ex-post* compliance assessment, systematic inspection, and third-party certification, following the Korean administrative law tradition; the label captures accountability exercised through *ex-post* administrative review.

Governance (*gwalli/geobeoneonseu*): Institutional and ministerial coordination, role definition, and accountability running upward through administrative tiers; we use the label in a managerial, hierarchical sense rather than merely referring to governance in general.

Public Good (*gonggongseong/gong-ik*): Collective welfare considerations and societal benefit beyond individual rights.

Harmony (*johwa/gyunhyeong*): Social balance, conflict minimization, and institutional coordination.

We label these Korean-salient in a comparative sense, not a proprietary or ‘essential’ one. They are also better read as long-standing features of Korean administrative practice than as recent imports, since collective welfare, oversight, and hierarchical accountability predate the current AI ethics vocabulary (Bell 2006; Pye and Pye 1985), unlike the explicit idiom of individual human rights, which is genuinely modern (Moyn 2010). None appears as a top-level principle in the consensus mappings (Fjeld et al. 2020; Jobin, Ienca, and Vayena 2019), yet all four are disproportionately salient here, and each aligns with a distinctive institutional form: not auditing in general but *ex-post* inspection by designated state bodies, not governance in general but ministerial coordination under hierarchical accountability.

Analytical Approach

Our analysis proceeds through six complementary steps, ordered as a pipeline: we (1) measure how prominently each principle features in each document, (2) validate the measure semantically, (3) compare prominence across document types to calculate translation gaps, (4) read the statutory text article by article to see whether a principle is named or made operative, (5) test robustness, and (6) estimate a regression of the joint effects. We report the semantic validation and robustness results in the supplementary appendix and describe each step's purpose and mechanics below.

Principle prominence. For each document, we calculate the frequency of principle keywords per 1,000 tokens: keyword density = (total keyword matches / document length in tokens) × 1,000. Keywords were assigned to principles in two steps: for the internationally-recognized set we translated the concept definitions in Fjeld et al. (2020) into their standard Korean policy equivalents, and for the Korean-salient set we derived terms inductively from the ethics guidelines, retaining a term only when it recurred across issuing ministries. Matching is on surface forms rather than stems, so recurring morphological variants (for example *gongjeong* and *gongjeongseong*, both rendering fairness, or *tumyeong* and *tumyeongseong* for transparency) are listed as separate entries, and terms whose dominant sense in statutory Korean is unrelated to the principle were excluded. Appendix A¹ lists every term and its source. Compared to manual coding, this approach is scalable; compared to LLM classification it is transparent and replicable (Grimmer and Stewart 2013; King, Lam, and Roberts 2017). Notably, keyword presence primarily measures discursive prominence, not necessarily full policy operationalization or robustness.

Semantic validation: To verify that keyword frequency differences reflect substantive normative content rather than incidental word usage, we employ two complementary validation methods. First, keyword-in-context (KWIC) extraction captures the 50 characters surrounding each keyword match, enabling qualitative verification that keywords appear in normatively relevant contexts. Second, we construct principle-level semantic networks using cosine similarity between keyword embeddings from a Korean-language transformer model (jhgan/ko-sroberta-multitask), visualized through network graphs and t-SNE projections. The resulting semantic network and co-occurrence structure are reported in Appendix F (Figures F1 and F2), together with the corpus composition by administration (Figure F3). These analyses address whether the same keyword carries a comparable normative meaning across document types.

¹ Supplementary materials (keyword dictionary, coding rules, and robustness results) are submitted as a separate appendix file and are also available at <https://osf.io/6wvk7/>

Translation gap calculation. The translation gap for a principle is its mean density in the ethics guidelines minus its mean density in the target set, which is either the non-binding policy documents or the binding law documents; positive values indicate erosion (gaps) and negative values amplification. Because a few topically specialized documents dominate some policy pathway means and skew distributions, we treat median-based tests as the primary analysis and report means alongside for comparability. Additional test statistics and effect sizes are reported in Appendix C.

Article-level operationalization coding. Operationalization is the extent to which a provision turns an abstract principle into concrete administrative or legal action. We apply this coding to the binding pathway only, since non-binding instruments do not typically create the obligations this coding approach centers. The statutory article is the unit of analysis. Each principle-article pair receives a binary score on five independent indicators: a binding obligation (the deontic form *hayeoya handa*, must, as against the hortatory *noryeokhayeoya*, shall endeavour), a named duty-bearer, a procedure (assessment, certification, inspection, record-keeping, or reporting), a penalty, and an individual remedy. Scores are aggregated to the share of a principle's articles meeting each indicator. We keep indicators separate rather than summing them into an index; a provision can for instance be operationalized partially, specifying who must do what and how, while carrying no penalty or remedy. Because Korean statutes place penalties in dedicated penalty (*beolchik*) and administrative-fine (*gwataeryo*) articles that cross-reference substantive provisions by number, enforcement was coded by resolving those references rather than by co-occurrence. Scoring applies explicit rules over statutory language rather than case-by-case judgment, so consistency derives from the rule set rather than from intercoder agreement; the full trigger expressions are listed in Appendix B. Finally, because a few documents dominate the non-binding pathway, where several principles have standard deviations exceeding their means, we repeat every ethics-to-target density comparison on medians using Mann-Whitney tests; the erosion pattern holds under both, with full results in Appendix D.

OLS regression. We first define the variables. 'Framework' is a principle's origin, internationally recognized (Fjeld et al. 2020) versus Korean-salient. 'Pathway' is the target instrument type, binding (statutes and enforcement decrees) versus non-binding (other policy documents). For simplicity, we abbreviate these in cases, describing them as the 'law' and 'policy' pathways. Finally, 'Administration' is the issuing government (Moon or Yoon) and year is mean-centered. We estimate translation gap ~ pathway + framework + administration + year to test whether the framework effect persists under controls; the four predictors explain only about 5% of the variance, so most of what shapes translation lies outside of the core variables we capture here.

Findings

Principle Prevalence and Translation Gaps

In ethics guidelines, our starting set of documents, the two principle sets are about equally prominent (7.8 vs 7.5 keywords per 1,000 tokens); under binding policy and law, they diverge sharply, the international set falling to 2.7 while the Korean-salient set holds at 7.4. As depicted in Table 1 and Figure 1, most internationally recognized principles erode in both pathways, and the binding pathway produces consistently larger gaps. Erosion is largest for justice/fairness (gap = 10.3, $d = 1.92$, $p_FDR < .001$), transparency (9.8, $d = 1.22$, $p_FDR = .002$), and non-maleficence (9.1, $d = 0.72$), followed by trust (6.5), freedom/autonomy (6.1), and responsibility (5.8). That is, the principles most emphasized in global ethics discourse erode most; normative prominence does not protect against it. Concretely, a fairness gap of 10.3 means fairness vocabulary occupies about ten fewer keyword slots per thousand words in statutes than in the ethics guidelines they ostensibly implement, against a baseline of 11.4.

Principle	Ethics	Policy	Law	Policy gap	Law gap
Justice/fairness (I)	11.38	3.29	1.12	8.1***	10.3***
Transparency (I)	11.25	6.58	1.45	4.7	9.8***
Non-maleficence (I)	21.23	18.19	12.14	3.0	9.1*
Trust (I)	8.06	7.52	1.55	0.5	6.5**
Freedom/autonomy (I)	6.44	2.37	0.35	4.1	6.1**
Responsibility (I)	7.08	2.68	1.26	4.4*	5.8**
Public good (K)	8.90	7.81	4.94	1.1	4.0
Dignity (I)	4.42	0.38	0.53	4.0	3.9
Privacy (I)	9.83	16.84	6.71	-7.0	3.1
Auditing/verif. (K)	11.66	17.84	9.10	-6.2	2.6
Sustainability (I)	3.24	4.05	1.96	-0.8	1.3
Beneficence (I)	1.73	1.71	0.47	0.0	1.3
Harmony (K)	0.75	1.01	0.57	-0.3	0.2
Solidarity (I)	1.54	1.73	2.50	-0.2	-1.0
Governance (K)	8.71	16.42	14.92	-7.7*	-6.2

Table 1: Principle prevalence and translation gaps. (I) = internationally recognized, (K) = Korean-salient. Densities are keyword matches per 1,000 tokens. Gap = ethics density minus target document density; positive values indicate erosion, negative values amplification. Asterisks on each gap denote Welch's t-tests against the ethics baseline: * $p < .05$, ** $p < .01$, *** $p < .001$.

International vs. Korean Principles

Figure 1 orders the fifteen principles by how much ground they lose on the way into binding law, and the ordering reflects our central result. Namely, the six largest losses all belong to the internationally recognized set (I): justice/fairness (law gap = 10.3), transparency (9.8), non-maleficence (9.1), trust (6.5), freedom/autonomy (6.1), and responsibility (5.8). The Korean-salient principles (K) cluster at the

other end. Their mean law gap is 0.13, statistically indistinguishable from zero, against an average 5.10 for the internationally recognized set (Table 2). The overall contrast is thus between erosion and retention rather than between erosion and amplification. The exceptions are nevertheless informative. Governance is the only principle with a negative gap in both pathways. Its amplification is robust in the non-binding policy pathway (mean gap = -7.7, median gap = -2.2, Mann-Whitney $p = .046$) though not under binding law, where the mean gap of -6.2 reverses to +1.1 on medians. Auditing/verification amplifies in the non-binding pathway (policy gap = -6.2) while showing a positive law gap (2.6), and privacy behaves the same way (policy gap = -7.0, law gap = 3.1) despite belonging to the internationally recognized set.

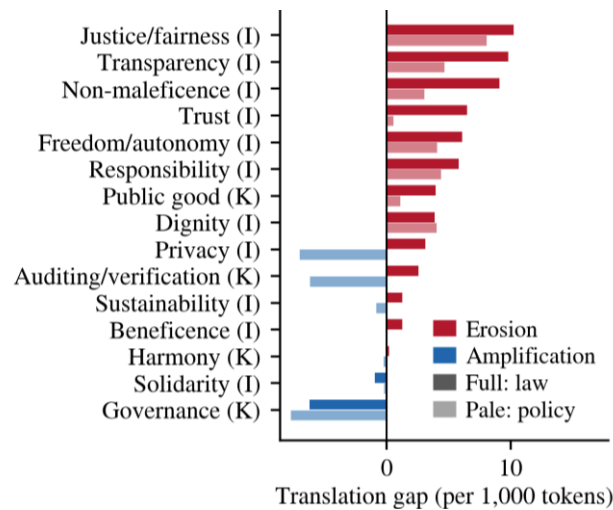


Figure 1: Translation gap by principle, relative to ethics guidelines. Red indicates erosion (positive gap), blue amplification (negative gap); solid bars are the binding ‘law’ pathway and pale bars the non-binding ‘policy’ pathway. (I) = internationally recognized, (K) = Korean-salient.

Privacy is the one rights-oriented principle with a dedicated enforcement body, the Personal Information Protection Commission established in 2020, which is what the demand-type account would predict. Harmony is flat throughout, and public good erodes (law gap = 4.0), like the internationally recognized set rather than the other Korean-salient principles. Because governance is often attached to administrative vocabulary, we checked its statutory contexts: of 853 governance keyword matches in the 19 binding instruments, 39.4% occur in an explicit AI, data, or digital context and 6.7% in an unambiguously generic administrative one, with the remainder in structural provisions such as committee and standard-setting articles (Appendix E). Auditing/verification also shows no significant difference from ethics guidelines in either pathway using medians; harmony is flat; and public good erodes.

Overall, internationally recognized principles show a mean binding law gap of 5.10 (SD = 3.74) and a mean policy gap of 1.90 (SD = 3.99). Meanwhile, domestically amplified principles show a mean law gap of 0.13 (SD = 4.50), statistically indistinguishable from zero, and a mean policy gap of -3.27 (SD = 4.33), indicating amplification. The asymmetry tracks institutional timing: principles requiring individual rights-enforcement depend on bodies built only recently (the National Human Rights Commission, 2001; the Personal Information Protection Commission, 2020), whereas oversight and audit draw on long-standing capacity.

Principle set (N)	Law mean	Law median	Policy mean	Policy median
International (11)	5.10	5.82	1.90	3.04
Korean-salient (4)	0.13	1.37	-3.27	-3.22

Table 2: Translation gaps by principle set and pathway. Cells are mean and median gaps in keywords per 1,000 tokens; positive values indicate erosion. Standard deviations are 3.74 and 3.99 for the internationally recognized set (law and policy) and 4.50 and 4.33 for the Korean-salient set.

Binding (Law) vs. Non-binding (Policy) Pathways

Ethics guidelines often present transparency and justice/fairness as entitlements owed to individuals, for instance framed around a right to know or protection from discriminatory treatment. Enacted laws reframe these principles as administrative requirements: data quality standards, system documentation obligations, and audit trail maintenance. The normative content shifts from protecting individual rights to managing institutional processes, arguably because the Korean administrative apparatus possesses established procedures for the latter but lacks mechanisms or deep understanding of the former.

The amplification of governance (law gap = -6.2, policy gap = -7.7) reveals another pattern: when principles require state-led management and oversight, the Korean bureaucracy does not merely translate them but actively expands them. Here, the developmental state apparatus treats AI governance as an extension of its existing management functions, applying established administrative hierarchies to a new technological domain. In short, on this interpretation which we cannot definitively show here, the principles that survive appear to be those the administrative apparatus is already organized to act on.

Finally, public good is the case where origin and demand come apart, which is why we treat its behavior as informative rather than anomalous. Its mean law gap is positive (4.0) while its median is negative (-2.7), so the direction is unstable and no consistent policy-pathway pattern appears. It is Korean-salient by origin, yet it names a diffuse collective good rather than an administrative routine, and it erodes like the international set.

Operationalization: Coverage, Not Drafting Gaps

One limitation of pure keyword-based analyses is that a principle can be named in a statute without having any substantive function or impact. To address this limitation in much of the extant translation literature, we score all binding statutes article by article. As noted previously, we suggest that two kinds of demand characterize the fifteen principles. Some ask the state to create or enforce entitlements that individuals hold against it or against firms: justice/fairness, transparency, privacy, dignity, and freedom/autonomy. Others ask the state to organize, monitor, and verify administrative activity: governance, auditing, and non-maleficence in its safety-assurance form. We refer to these as rights-oriented and management-oriented demands, and we suggest the distinction matters because the two draw on different state capacities.

We find that, within the 19 binding law instruments, principles that appear at all are usually written in operative terms, and rights-oriented principles are no exception: transparency carries an obligation in 79.7% of the articles that invoke it and justice/fairness in 70.3%, against 59.9% for governance and 57.5% for auditing. Table 3 gives the number of statutory articles in which each appears and the percentage of those articles that impose an obligation, name a duty-bearer, set out a procedure, attach a penalty, and provide a remedy. Every principle appears somewhere in the binding instruments, from 327 articles for non-maleficence down to 17 for freedom/autonomy, so the differences are of breadth rather than of presence.

Principle	Arts.	Obl.	Duty	Proc.	Pen.	Rem.
Non-maleficence (I)	327	61.5	87.2	23.5	10.7	10.7
Governance (K)	267	59.9	90.6	21.3	10.1	9.4
Privacy (I)	199	55.8	85.9	17.6	5.5	14.1
Audit./verif. (K)	179	57.5	91.1	47.5	17.3	7.3
Public good (K)	153	63.4	91.5	19.6	2.0	10.5
Sustainability (I)	102	55.9	90.2	16.7	2.9	7.8
Solidarity (I)	84	66.7	95.2	20.2	1.2	4.8
Transparency (I)	74	79.7	86.5	29.7	8.1	17.6
Responsibility (I)	72	77.8	81.9	23.6	9.7	15.3
Trust (I)	65	69.2	89.2	21.5	6.2	1.5
Beneficence (I)	44	65.9	79.5	27.3	15.9	20.5
Justice/fairness (I)	37	70.3	75.7	24.3	5.4	8.1
Harmony (K)	30	73.3	80.0	10.0	10.0	0.0
Dignity (I)	22	72.7	72.7	18.2	4.5	9.1
Freedom/auto. (I)	17	64.7	88.2	23.5	5.9	11.8
All international	1043	64.0	86.6	21.9	7.5	11.1
All Korean	629	60.7	90.5	27.8	10.2	8.6

Table 3: Article-level operationalization coding, all fifteen principles. (I) = internationally recognized, (K) = Korean-salient. Arts. = number of statutory articles invoking the principle. The remaining columns give the percentage of those articles that carry a binding obligation (Obl.), a named duty-bearer (Duty), a procedural requirement (Proc.), a penalty (Pen.), and an individual remedy (Rem.). Counts are principle-article pairs; an article mentioning more than one principle is counted under each.

Overall, the asymmetry lies in coverage, not depth. Justice/fairness appears in 37 articles (column Arts.) across 12 of 19 instruments and freedom/autonomy in 17 across 4. Governance appears in 267 articles in all 19 binding laws, and auditing in 179 across 17. Where the legislature addresses fairness or transparency it writes them in operative terms: 70.3% and 79.7% of those articles impose a binding obligation, above the corpus average of 64.0%. Auditing is the most procedurally dense principle, with 47.5% of its articles specifying an assessment, certification, or inspection step, and the most likely to carry a penalty cross-reference (17.3%). Also of note, the resulting duties are typically owed to a supervising ministry rather than to an affected person.

In sum, the ethics-to-law gap may be better described as defined by varying coverage breadth rather than drafting quality or discursive ethics-washing, a result that cuts across the origin of principles. Yet, the findings also suggest that it is possible to operationalize all of the principles; leaving open why some receive less coverage in the first place.

Regression Summary

Next, we regress the translation gap measure on pathway + framework + administration + year to test whether the International-versus-Korean patterns identified above persist under controls. Associational results from an OLS model estimated on principle-document pairs are presented in Table D1, Appendix D. In brief, we find that framework (principle origin) and pathway (binding or non-binding) are reliable but small factors, while most of what shapes translation lies outside the four predictors. The model explains only about 5% of the variance ($R^2 = .053$, $F(4, 685) = 9.57$, $p < .001$), suggesting much more research is needed to better characterize, model, and explain the translation gap, though we offer this analysis approach as a preliminary contribution.

In terms of effect size in our associations, framework (Korean or International) carries the largest coefficient ($b = -.509$), and pathway confirms the binding-pathway amplification pattern ($b = -3.21$): the non-binding pathway is associated with more principle translation than the binding one. Neither administration ($b = 0.40$, $p = .81$) nor year ($b = -0.11$, $p = .79$) approaches significance. Finally, because each document contributes fifteen rows, we re-estimated the model with standard errors clustered by document; the significance of framework and pathway are unchanged.

Administration Comparison

Figure 2 presents the administration-level findings, comparing differences across two administrations (and associated political-ideological orientations). On means, the gap appears to narrow sharply (law 6.18 to 2.11; policy 1.63 to -0.73), but the narrowing does not survive scrutiny: on medians the law gap moves only from 3.68 to 3.02, the policy gap does not narrow at all, and the administration coefficient is null. On closer inspection, the apparent convergence comes from a few topically concentrated Yoon-era policy documents.

For Korean-salient principles, governance amplification is consistent across both administrations (blue for Moon; red for Yoon). Meanwhile, auditing shows a striking shift: the Moon policy gap of 0.5 (straight translation) versus the Yoon policy gap of -7.5 (strong amplification), may reflect the Yoon administration's greater emphasis on AI safety evaluation and the implementation of the AI Basic Act's auditing provisions.

Yet, Figure 2 reveals that this aggregate convergence also masks additional principle-level variation. Justice/fairness erosion persists across both administrations (Moon law gap = 12.3, Yoon = 9.6). Privacy reverses across administrations: it was underrepresented in Moon-era law (gap = 5.0), then proliferated in Yoon-era policy documents (gap = -5.8) after the Personal Information Protection Commission was established as an independent regulator. This micro-level pattern is consistent with the institutional mechanisms argument advanced here: when the state creates a dedicated body capable of rights-enforcement (PIPC for privacy), even principles requiring individual protection can translate successfully. The barrier is not the principle's normative content but the absence of institutional infrastructure to deliver it.

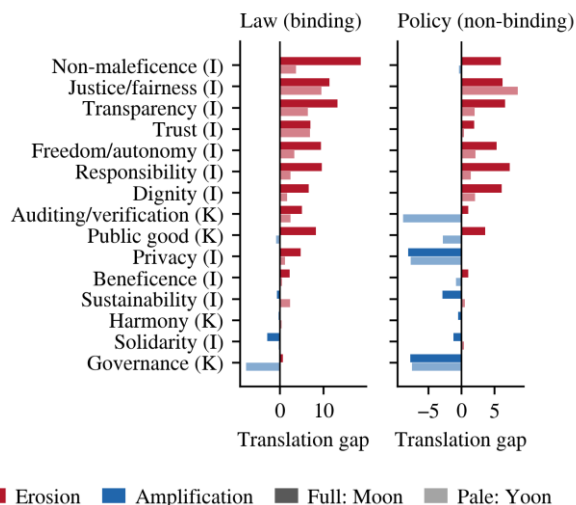


Figure 2: Translation gaps by administration and pathway. Red indicates erosion, blue amplification; solid bars are the Moon administration and pale bars the Yoon administration.

We had expected institutional actors to accumulate experience and become ‘better’ at translation, and the law pathway is partly consistent with that (justice/fairness erosion falls from 11.4 to 9.5, transparency from 13.2 to 6.4). But neither measures of political administration nor year predicts translation outcomes once framework and pathway are accounted for. One reading of this null result is that AI policy in Korea has not yet acquired a clear partisan cleavage: comparative work finds AI policy in the United States started in a bipartisan fashion and is only gradually polarizing over time as AI becomes characterized by more divisive policy tools or framings (Jacobson, Schiff, and Schiff 2026). More over-time study will be needed to assess this.

Discussion and Implications

State Capacity and Selective Translation

Why does a state with strong bureaucratic capacity and an explicit commitment to AI ethics still produce rules that privilege administrative coordination over individual rights? Our findings indicate that, because the two principle sets are equally prominent in ethics guidelines, the divergence appears only at codification. The design is observational, so causal claims are limited, though the pattern is broadly consistent across the 15 principles, both pathways, both administrations, and the 69 documents. Of course, institutional fit, the theory emphasized here, is not the only candidate explanation: our preliminary OLS model explains only about 5% of the variance in translation gaps. Political priorities, strategic signaling, international competition, industry lobbying (Korean AI development is concentrated in a few large firms with established channels into ministries), and strategic efforts by policymakers to leave rights provisions more flexible could each contribute.

Yet one plausible and understudied factor concerns the fit between what principles demand of the state and what it can deliver. The multiple streams literature separates two dimensions of that fit, technical feasibility and value acceptability (Kingdon 2011; Schiff 2023), on which our evidence bears unequally. Feasibility is observable through institutional dating: principles requiring rights-enforcement machinery depend on bodies created in 2001 and 2020, whereas oversight and verification extend routines that long predate AI, and the ordering of principle erosion follows that temporal difference. We do not measure value acceptability directly outside of what can be inferred from our core documents, so we treat it as a conjecture that this design cannot adjudicate. We suggest that additional analysis involving legislative records, elite interviews, and culturally-sensitive research methods would be needed to examine this worthwhile question. Testing it would require different evidence: legislative and committee records, elite interviews, or, for instance, a survey experiment with legislative staffers that varies how a principle is framed. However, our results do indicate that international principles appear equally in normative Korean documents, which suggests some degree of shared interest or normative high-level acceptability, leaving technical feasibility the stronger explanation.

Regarding technical and institutional feasibility (or policy capacity) then, we argue that because the Korean bureaucracy has decades of experience with ministerial coordination, compliance monitoring, and sectoral verification (Ginsburg 2003; Im et al. 2014), it readily extends existing capacity to a new domain, whereas justice/fairness, transparency, and autonomy demand capacities with less precedent in Korean administrative law. The asymmetry is observable within the internationally recognized set as well as between frameworks: privacy, alone among rights-oriented principles, is not eroded in the policy pathway relative to the others, and it is also the one domain where Korea has operated a dedicated enforcement body since 2020.

Pathway-Dependent Translation

The binding pathway produces larger translation gaps than the non-binding one ($b = -3.21$). To explain this, we suggest that binding force imposes a drafting discipline that filters normative content: an enforceable provision may need to name a duty-bearer, specify conduct, and attach a consequence, and principles that resist that form appear to not to be written at all. Additionally, because (legislative) statutes and (executive) enforcement decrees behave alike in our data, this explanation for translation success or failure appears to reflect legal force rather than authorship source.

Plans and strategies, by contrast, accommodate broader normative language. Legal codification, often treated as the gold standard of implementation, therefore erodes normative content more than non-binding policy does, so studies confined to legislation may overstate translation failure. Along these lines, our corpus also excludes procurement, standards, and certification, which can translate principles that even formal law leaves abstract (Ben Dor and Coglianese 2021); in Korea, for instance, the AI Basic Act's high-impact provisions are being given content through ministerial guidelines and a voluntary certification scheme. Rights-oriented content may therefore be relocating into instruments such as standards, certification, and audit schemes rather than disappearing. Future scholarship should aim to conceptualize and measure different translation pathways beyond the ones we have addressed here, noting regional and institutional specificities in how laws, regulations, branches of government, and so on function.

Theoretical and Empirical Contributions

These findings advance technology governance theory in three ways. First, they challenge a universalist approach to AI ethics. International initiatives such as the OECD AI Principles and the EU AI Act emphasize comprehensive principle coverage, assuming normative consensus will translate into implementation. Our findings complicate this assumption: effective governance design requires shifting from “which principles should we implement?” to “what state capacities must we build to implement them?”

Second, we contribute to research on policy design by separating pathways by legal force, showing that instruments differing in bindingness produce different translation outcomes, so studies reporting a single ethics-to-policy gap may be averaging over that difference. Building on prior work that examines enacted or proposed law (Lee and Lee 2025), we also analyze the executive policy documents through which administrations turn principles into action plans, and compare the two pathways directly, empirically.

Third, our broader contribution is a preliminary though replicable framework: by defining two translation pathways, semantically validating principle differences, measuring translation gaps through keyword density and article-level coding of provisions, this design can be applied across countries, administrations, and topics using materials governments already publish. This framework merits critique and expansion as well, in terms of conceptualization and measurement: our regression analysis suggests as much.

Policy Implications

We highlight three policy implications. First, the erosion of justice/fairness and transparency suggests the ethical values that scholars, advocates, and other stakeholders hold dear may need dedicated institutional infrastructure, such as designated oversight bodies and mandatory algorithmic impact assessments, to survive codification into binding law.

Second, and inversely, the retention of governance and auditing in Korean policy shows that AI governance can build on existing institutional strengths. States with strong audit traditions may find leveraging them more effective than importing rights-based frameworks wholesale. This is a question of policy capacity (Wu, Ramesh, and Howlett 2015): a state with deep inspection capacity and shallow rights machinery will produce auditable AI, while one with strong courts and weak administrative reach would fail differently. We propose a sequencing approach: begin where capacities align and build the missing ones in parallel.

Third, and against our expectations, gaps did not narrow across administrations. Successive governance cycles may not be sufficient, because policy learning runs alongside path dependency: the very grooves that make governance language easy to write also make aspirational ethics reform hard. Path dependencies may need to be actively countered.

Limitations

Several limitations bound our arguments. Keywords captures discursive prominence rather than effect, the article-level enforcement coding covers one pathway only, and longer ethics guidelines have lower keyword density ($r = -.53$), which is why we do not rest the administration comparison on means. Finally, all the methodologies we apply are observational and do not involve process-tracing, so unobserved causes and reverse causation cannot be excluded.

Conclusion

Using 69 Korean AI governance documents spanning two administrations, we find that internationally recognized principles erode during translation while Korean ethical principles, including those embedded in Korean governance and bureaucracy, are retained or amplified. We further find that the binding pathway imposes stricter normative filters than the non-binding one. The Korean case carries a broader lesson: normative consensus at the ethics level does not guarantee implementation at the governance level. The principles that survive translation are those invoking functions the government already performs, not necessarily those with the strongest normative justification, greatest consensus, or even largest descriptive presence. Closing the ethics-governance gap may require constructing the institutional infrastructure or policy capacity to sustain principles through the translation process. The question facing AI governance globally is less whether the right principles have been identified than whether states will build the machinery those principles require in order to realize the aims of AI ethics.

Ethics Statement

This study analyzes publicly available government documents and involves no human subjects or personal data. The principle dictionary encodes a particular reading of Korean policy vocabulary, and our classification of principles by the demand they place on the state is an analytic choice rather than a neutral description; both are documented in the supplementary appendix so that readers can assess them. Finally, we caution against reading the finding that rights-oriented principles translate poorly as a normative endorsement of that outcome.

References

- Amsden, A. H. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.
- Bareis, J. and Katzenbach, C. 2022. Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics. *Science, Technology, & Human Values* 47(5): 855–881.
- Baumgartner, F. R.; Berry, J. M.; Hojnacki, M.; Kimball, D. C.; and Leech, B. L. 2009. *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago: University of Chicago Press.
- Bell, D. A. 2006. *Beyond Liberal Democracy: Political Thinking for an East Asian Context*. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400827466>.
- Bell, D. A. 2010. *China's New Confucianism: Politics and Everyday Life in a Changing Society*. Princeton: Princeton University Press.
- Ben Dor, L. M. and Coglianese, C. 2021. Procurement as AI Governance. *IEEE Transactions on Technology and Society* 2(4): 192–199.
- Calo, R. 2017. Artificial Intelligence Policy: A Primer and Roadmap. *UC Davis Law Review*, 51: 399–435.
- Carpenter, D. 2002. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862–1928*. Princeton: Princeton University Press. <https://doi.org/10.1515/9780691214078>.
- Cohen, J. 2013. *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge.
- Corrêa, N. K.; Galvão, C.; Santos, J. W.; Del Pino, C.; Pinto, E. P.; Barbosa, C.; et al. 2023. Worldwide AI ethics: A Review of 200 Guidelines and Recommendations for AI Governance. *Patterns* 4(10): 100857.
- DiMaggio, P. J. and Powell, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48(2): 147–160.
- Djeffal, C.; Siewert, M. B.; and Wurster, S. 2022. Role of the state and responsibility in governing artificial intelligence: a comparative analysis of AI strategies. *Journal of European Public Policy* 29(11): 1799–1821. <https://doi.org/10.1080/13501763.2022.2094987>.
- Eckstein, H. 2000. Case Study and Theory in Political Science. In *Case Study Method*, 119–164, London: Sage Publications.
- Entman, R. M. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43(4): 51–58.
- Evans, P. B. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton, NJ: Princeton University Press.
- Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; and Srikumar, M. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Research Publication*, (2020–1).
- Floridi, L. 2018. Soft Ethics, the Governance of the Digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376(2133): 20180081. <https://doi.org/10.1098/rsta.2018.0081>.
- Gerring, J. 2006. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Ghimire, A. and Edwards, J. 2024. From Guidelines to Governance: A Study of AI Policies in Education. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, 299–307, Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64312-5_36.
- Ginsburg, T. 2003. *Judicial Review in New Democracies: Constitutional Courts in Asian Cases*. Cambridge: Cambridge University Press.
- Grimmer, J. and Stewart, B. M. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political analysis* 21(3): 267–297.
- Haas, P. M. 1992. Introduction: Epistemic Communities and International Policy Coordination. *International Organization* 46(1): 1–35.
- Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30(1): 99–120. <https://doi.org/10.1007/s11023-020-09517-8>.
- Hildebrandt, M. 2020. *Law for Computer Scientists and Other Folk*. Oxford: Oxford University Press.
- Im, H. B. 2004. Faltering Democratic Consolidation in South Korea: Democracy at the End of the ‘Three Kims’ Era. *Democratization* 11(5): 179–198. <https://doi.org/10.1080/13510340412331304642>.
- Im, T.; Cho, W.; Porumbescu, G.; and Park, J. 2014. Internet, Trust in Government, and Citizen Compliance. *Journal of Public Administration Research and Theory* 24(3): 741–763.
- Jacobson, R.; Schiff, D. S.; and Schiff, K. J. 2026. How New Issues Become Polarized: Partisan Triggers and Subsystem Shopping in Early AI Policymaking. *Policy & Internet* 18(1): e70020.
- Jasanoff, S. 1998. *The Fifth Branch: Science Advisers as Policymakers*. Cambridge, MA: Harvard University Press.
- Jasanoff, S. 2003. Technologies of Humility: Citizen Participation in Governing Science. *Minerva* 41(3): 223–244. <https://doi.org/10.1023/A:1025557512320>.
- Jobin, A.; Ienca, M.; and Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1(9): 389–399.
- Jørgensen, T. B.; and Bozeman, B. 2007. Jørgensen, T. B.; and Bozeman, B. 2007. Public Values: An Inventory. *Administration & Society* 39(3): 354–381.
- Kaminski, M. E. 2023. Regulating the Risks of AI. *Boston University Law Review*, 103: 1347.

- Kaminski, M. E. and Malgieri, G. 2021. Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations. *International Data Privacy Law*11(2): 125–144.
- King, G.; Lam, P.; and Roberts, M. E. 2017. Computer-Assisted Keyword and Document Set Discovery from Unstructured Text. *American Journal of Political Science*61(4): 971–988. <https://doi.org/10.1111/ajps.12291>.
- Kingdon, J. W. 2011. *Agendas, Alternatives, and Public Policies*, 2nd ed. Boston: Longman.
- Larsson, S. 2020. On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*7(3): 437–451.
- Lee, J. W. and Lee, K. 2025. Building a Consensus: Harmonizing AI Ethical Guidelines and Legal Frameworks in Korea for Enhanced Governance. *Government Information Quarterly*42(3): 102060.
- Mittelstadt, B. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature machine intelligence*1(11): 501–507.
- Mohamed, S.; Png, M.-T.; and Isaac, W. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology* 33: 659–684.
- Moyn, S. 2010. *The Last Utopia: Human Rights in History*. Cambridge, MA: Belknap Press of Harvard University Press.
- Mulligan, D. K.; Kroll, J. A.; Kohli, N.; and Wong, R. Y. 2019. This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology. Proceedings of the ACM on Human-Computer Interaction 3(CSCW): 1–36.
- OECD. 2019. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. Paris: OECD.
- Png, M.-T. 2022. At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 1434–1445.
- Pye, L. W. and Pye, M. W. 1985. *Asian power and politics: The cultural dimensions of authority*. Cambridge, MA: Harvard University Press.
- Reisman, D.; Schultz, J.; Crawford, K.; and Whittaker, M. 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency. NIST. <https://www.nist.gov/system/files/documents/2021/10/04/aiarepor t2018.pdf>.
- Sabatier, P. A., ed. 2007. *Theories of the Policy Process*, 2nd ed. Boulder, CO: Westview.
- Schiff, D. S. 2023. Looking through a Policy Window with Tinted Glasses: Setting the Agenda for U.S. AI Policy. *Review of Policy Research* 40(5): 729–756.
- Schiff, D. S.; Laas, K.; Biddle, J. B.; and Borenstein, J. 2022. Global AI Ethics Documents: What They Reveal About Motivations, Practices, and Policies. In *Codes of Ethics and Ethical Guidelines*, 121–143, Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-86201-5_7.
- Schiff, D.; Borenstein, J.; Biddle, J.; and Laas, K. 2021. AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection. *IEEE Transactions on Technology and Society*2(1): 31–42.
- Stone, D. A. 1989. Causal Stories and the Formation of Policy Agendas. *Political Science Quarterly* 104(2): 281–300.
- Sunstein, C. R. 1995. Incompletely Theorized Agreements. *Harvard Law Review* 108(7): 1733–1772.
- UNESCO. 2021. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.
- Veale, M. and Borgesius, F. Z. 2021. Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*22(4): 97–112. <https://doi.org/10.9785/crl-2021-220402>.
- Wade, R. 2018. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton: Princeton University Press.
- Whittlestone, J.; Nyrup, R.; Alexandrova, A.; and Cave, S. 2019. The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200, Honolulu HI USA: ACM. <https://doi.org/10.1145/3306618.3314289>.
- Wong, P.-H. 2020. Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI. *Philosophy & Technology*33(4): 705–715. <https://doi.org/10.1007/s13347-020-00413-8>.
- Woo-Cumings, M. 2019. *The Developmental State*. Ithaca, NY: Cornell University Press.
- Wu, X.; Ramesh, M.; and Howlett, M. 2015. *Policy Capacity: A Conceptual Framework for Understanding Policy Competences and Capabilities*. *Policy and Society* 34(3–4): 165–171.
- Zhu, M. and Xu, C. 2023. International Soft Law Governance of Artificial Intelligence Ethics: Current Situation, Challenges and Countermeasures. *Bulletin of Chinese Academy of Sciences (Chinese Version)*38(7): 1037–1049.
- ÓÉigeartaigh, S. S.; Whittlestone, J.; Liu, Y.; Zeng, Y.; and Liu, Z. 2020. Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance. *Philosophy & Technology*33(4): 571–593. <https://doi.org/10.1007/s13347-020-00402-x>.