Animal Welfare and Policy Risk Index (AWPRI): Constructing and Validating a Cross-National Governance Risk Measure, 25 Countries, 2004–2022

Jason Hung

Futurekind AI Fellow, Electric Sheep

Abstract

While animal welfare governance continues to be influenced by technological advancement and automation, there is an absence of longitudinal, cross-country, quantitative index that simultaneously features the governance baseline, the direction of policy change, and the compounding risk posed by agricultural artificial intelligence (AI) adoption. This paper introduces the Animal Welfare and Policy Risk Index (AWPRI), a composite risk index covering 25 countries over the period 2004–2022 (N= 475 country-year observations). The AWPRI is constructed from 15 variables organised across three equal-weighted conceptual layers: Current Welfare State (L1), Policy Trajectory (L2), and AI Amplification Risk (L3). Variables are normalised to [0, 1] using min-max scaling, with higher values denoting greater policy risk. The index is validated through k-means cluster analysis (k= 4; silhouette coefficient = 0.447), principal component analysis (PCA) of the 15-variable cross-section, and sensitivity analysis under ± 10 percentage-point layer weight perturbation (mean Spearman ρ = 0.993, minimum 0.979; mean Adjusted Rand Index (ARI) = 0.684, range 0.477–1.000). Our Hausman specification test favours random-effects (RE) panel estimation (H= 2.55, p= 0.467). We use a difference-in-differences (DiD) design to exploit the 2019 AI governance risk classification divergence and find that countries identified as high-AI-governance-risk carry AWPRI scores 0.080 points higher than their low-risk counterparts, after controlling for country and year fixed effects (β= 0.080, SE= 0.005, p< 0.001). The L3 layer records the highest mean score in the 2022 cross-section (0.552, SD = 0.175), significantly exceeding both L1 (Wilcoxon W= 102,651, p< 0.001) and L2 (W= 99,295, p< 0.001). China (0.802), Vietnam (0.612), and Thailand (0.586) record the highest composite risk scores in 2022; the United Kingdom (0.308) the lowest. AutoRegressive Integrated Moving Average (ARIMA)-based projections indicate that Thailand, Brazil, and Argentina face AWPRI risk deterioration by 2030. The AWPRI and its interactive visualisation are publicly accessible at https://awpri-dashboard.streamlit.app.

Keywords*: animal welfare policy; artificial intelligence; precision livestock farming; composite index; governance gap; difference-in-differences; panel data*

1. Introduction

Approximately 80 billion land animals are slaughtered annually within global food systems [1]. The scale of this figure renders the institutional underinvestment in animal welfare governance both empirically significant and policy-relevant. Comparative political science and public policy scholarship have been slow to develop quantitative frameworks for cross-country welfare governance assessment. Where animal welfare is scholarly discussed, analysis is predominantly shaped by normative or legal terms [2, 3], or confined to case studies of discrete regulatory regimes [4]. There is an absence of a longitudinal, cross-country, quantitative instrument that tracks how animal welfare governance performs over time, whether those trajectories are improving or deteriorating, and whether emerging technological forces compound pre-existing governance gaps.

The rapid commercialisation of artificial intelligence (AI) in livestock production makes the discussion of technologically facilitated animal welfare regulation increasingly timely and relevant. Computer vision systems for automated lameness detection, AI-driven feed optimisation, and predictive disease modelling are now commercially deployed across major livestock-producing economies [7, 8]. Market projections for the precision livestock farming (PLF) sector reach USD 19.87 billion by 2032 [9]. A body of literature argues that AI enables earlier detection of welfare problems and reduces reliance on invasive interventions [10, 11]. Additional literature raises substantive concerns that PLF's welfare-positive claims remain unproven at commercial scale, and that AI-driven intensification poses systemic threats to welfare in jurisdictions whose regulatory frameworks were not designed to address algorithmic accountability [12, 13].

Existing composite measures of animal welfare governance, including the World Animal Protection's Animal Protection Index [5] and Hårstad's scoping review [3], feature legislative text at a single point in time. Neither instrument tracks law enforcement dynamics, policy reform trajectories, or the compounding effect of technological adoption on governance gaps. This paper addresses these limitations through the introduction and analysis of the Animal Welfare and Policy Risk Index (AWPRI).

1.1 Research Questions and Contributions

This paper pursues three research questions. First, how can animal welfare policy risk be operationalised as a measurable, cross-country comparable composite index sensitive to the governance implications of AI adoption in agriculture? Second, what patterns of risk distribution, clustering, and temporal change emerge across 25 countries between 2004 and 2022? Third, how does AI adoption in agriculture interact with pre-existing governance conditions and trajectories, and what national risk profiles are projected to emerge by 2030?

The paper makes four contributions.

1. It introduces the AWPRI: the first longitudinal, cross-country, AI-sensitive composite risk index for animal welfare governance, covering 25 countries over 19 years.

2. It validates the index through k-means cluster analysis, principal component analysis (PCA), Hausman specification testing, and a sensitivity analysis under layer weight perturbation.
3. It employs a difference-in-differences (DiD) design to estimate the effect of AI governance risk classification divergence on AWPRI trajectories, providing the first quasi-experimental evidence linking AI governance status to animal welfare policy risk.
4. It presents AutoRegressive Integrated Moving Average (ARIMA)-based projections to 2030 for all 25 countries with 95% confidence intervals, identifying which national risk profiles are projected to deteriorate absent policy intervention.

2. Related Work

2.1 Animal Welfare Governance and Composite Indices

Composite indices are well-established instruments for cross-country governance comparison. The Human Development Index [15], the Environmental Performance Index [16], and the Global Peace Index [17] demonstrate that multidimensional governance circumstances can be reduced to measurable scores while retaining policy interpretability. In the animal welfare domain, Browning [18] argues explicitly that multidimensional welfare measurement frameworks can support policy analysis. However, to date, no composite index applies such a framework to the intersection of animal welfare governance and AI adoption.

The World Animal Protection's Animal Protection Index [5] rates 50 countries on legislative capacity at a single point in time. Hårstad's scoping review of farm animal welfare governance [3] similarly prioritises legislative text and political drivers. He finds that policy change is neither linear nor easily predictable. Neither instrument takes into account the enforcement dynamics, temporal trajectories, or the risk introduced by AI-driven agricultural intensification. The Animal Law Foundation [6] has documented that fewer than 2.5% of farms in England were inspected in 2024, with 19% of inspected farms found in breach of welfare laws and fewer than 1% of violations resulting in prosecution. This enforcement gap illustrates the dimension that static legislative indices fail to highlight.

2.2 AI and PLF

Tuyttens et al. [12] identify 12 welfare threats specific to PLF adoption, including the displacement of human observation by algorithms, the intensification of stock enabled by automated monitoring, and the commercial incentives to use AI for productivity maximisation over welfare improvement. These concerns are amplified in jurisdictions with weak baseline welfare legislation, in which AI adoption can accelerate production intensification without triggering comparable regulatory responses [14]. Elliott and Werkheiser [13] argue that existing PLF transparency frameworks remain conceptually underdeveloped, and that most AI agricultural systems operate without welfare-specific accountability mechanisms. The AWPRI's L3 layer is designed specifically to quantify the compounding risk associated with this governance-technology asymmetry.

2.3 Panel Data Approaches to Governance Measurement

Fixed-effects (FE) and random-effects (RE) panel regression are standard approaches to exploiting longitudinal cross-country differences in governance indices. Hausman [19] specification tests are the conventional criterion for model selection. A significant Hausman statistic indicates that country-specific effects are correlated with the regressors, favouring the FE estimator, while a non-significant outcome renders the RE estimator valid and more efficient. DiD designs have been applied to identify causal effects of policy interventions in governance research [20]. Our statistical analysis provides the first quasi-experimental evidence in the animal welfare governance literature.

3. Methods

3.1 Country and Time Period

The AWPRI panel dataset covers 25 countries across six global regions over 19 years (2004−2022), resulting in a total of (25x19=) 475 country-year observations. Countries were selected to maximise regional diversity, data availability, and change in welfare legislative capacity, incorporating the world's largest livestock producers, the most progressive welfare legislatures, and major emerging economies undergoing rapid agricultural AI adoption. The 25 countries featured in this study are: Argentina, Australia, Brazil, Canada, China, Denmark, France, Germany, India, Italy, Japan, Kenya, Mexico, the Netherlands, New Zealand, Nigeria, Poland, South Africa, South Korea, Spain, Sweden, Thailand, the United Kingdom, the United States, and Vietnam. The full dataset is accessible at https://awpri-dashboard.streamlit.app. All analyses reported in this paper use the *panel_awpri_normalized.csv* dataset, which is the identical source used for data visualisation in the AWPRI interactive dashboard.

3.2 Variable Selection and Operationalisation

Fifteen variables are assigned across three equal-weighted conceptual layers, with five variables per layer. Layer 1 (L1: Current Welfare State) measures the governance baseline: (1) animal rights legislative framework; (2) rule of law index (risk-coded); (3) farmed animals per capita; (4) aquaculture share of production; and (5) meat consumption per capita. Layer 2 (L2: Policy Trajectory) features the direction and pace of governance change: (6) animal rights trend score (year-on-year legislative change); (7) plant protein risk; (8) civic space risk; (9) civil liberties risk; and (10) public concern proxy. Layer 3 (L3: AI Amplification Risk) quantifies the compounding effect of AI adoption in agriculture: (11) AI governance risk; (12) AI welfare research alignment; (13) AI sentience research risk; (14) specialist bias ratio in AI systems; and (15) livestock AI patent intensity. All 15 variables are coded such that higher values represent greater policy risk. Data are drawn from the World Animal Protection's Animal Protection Index [5], FAO FAOSTAT [1], the V-Dem Democracy Index [22], the Oxford Insights Government AI Readiness Index [23], the Stanford AI Index [24], and patent databases via OpenAlex. Missing values (approximately 7.3% of observations) are imputed using linear interpolation within country time series.

3.3 AWPRI Construction

All 15 variables are normalised to [0, 1] using min-max normalisation across the full 2004−2022 panel, enabling valid cross-country and cross-temporal comparison. Each layer score is the unweighted mean of its five constituent variables:

$$L_{1it} = (1/5) \sum \{k \in \mathcal{K}_1\} \, v_{kit}$$

$$L_{2it} = (1/5) \sum \{k \in \mathcal{K}_2\} \, v_{kit}$$

$$L_{3it} = (1/5) \sum \{k \in \mathcal{K}_3\} \, v_{kit}$$

where $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{K}_3$ denote the five-variable indicator sets for Layer 1, Layer 2, and Layer 3, respectively, as defined in Table A1 (Appendix), and where v_{kit} denotes the normalised value of variable k for country i in year t.

The composite AWPRI score is the unweighted mean of the three layer scores:

$$AWPRI_{it} = (L_{1it} + L_{2it} + L_{3it} \,/\, 3$$

Equal weighting is applied following Organisation for Economic Co-operation and Development (OECD) and Joint Research Centre (JRC) recommendations for composite indicators when no strong prior evidence exists for differential weighting across dimensions [21]. The robustness of this decision is evaluated through a sensitivity analysis described in Section 3.7.

3.4 Cluster Analysis and Validation

A four-tier risk typology is defined using score-based thresholds: Critical (≥ 0.55), High (0.45–0.55), Moderate (0.35–0.45), and Low (< 0.35). These boundaries are validated using k-means cluster analysis on the 2022 cross-section of composite and layer scores, with the optimal k determined through the elbow method and silhouette coefficient analysis. Cluster robustness is evaluated using three complementary metrics, namely (1) the silhouette coefficient, (2) the Calinski–Harabasz index, and (3) the Davies–Bouldin index.

3.5 Forecasting

Country-level AWPRI trajectories are projected to 2030 using ARIMA models estimated separately for each country, with model order selection via Akaike Information Criterion (AIC) minimisation. Forecast uncertainty is represented by 95% confidence intervals. All models are implemented in Python using the statsmodels library.

3.6 Statistical Analysis

A total of seven complementary inferential analyses are conducted.

First, Wilcoxon signed-rank tests are used to evaluate whether L3 scores are systematically higher than L1 and L2, both across the full panel (N $=$ 475) and in the 2022 cross-section (n $=$ 25). The signed-rank test is preferred over the parametric t-test given the bounded, non-normal distribution of normalised layer scores.

Second, a Spearman rank correlation matrix is computed for the 15 constituent variables on the 2022 cross-section to assess construct validity and detect potential multicollinearity in the index structure.

Third, a Kruskal–Wallis test followed by pairwise Mann–Whitney U tests with Bonferroni correction are applied to test whether AWPRI scores differ significantly across risk tiers.

Fourth, a Hausman specification test is conducted to choose between FE and RE panel estimators.

The test statistic is:

$$H = (\beta^{FE} - \beta^{RE})^T \ [Var(\beta^{FE}) - Var(\beta^{RE})]^{-1} \ (\beta^{FE} - \beta^{RE}) \sim \chi^2(K)$$

where β^{FE} and β^{RE} denote FE and RE coefficient vectors, respectively, and K is the number of time-varying regressors. A significant statistic ($p < 0.05$) implies a systematic difference between the estimators, favouring FE.

Fifth, a DiD design exploits the divergence in country-level AI governance risk classification that emerged from the 2019 Oxford Insights Government AI Readiness Index. Countries are classified as treated ($ai\_governance\_risk = 1.0$ in 2019, n = 14) and control ($ai\_governance\_risk = 0.0$, n = 11). The pre-period covers 2004–2016; the post-period refers to 2019–2022, omitting the 2017–2018 transition years. The estimating equation is:

$$AWPRI_{it} = \alpha + \beta(Post_t \times Treat_i) + \gamma_i + \delta_t + \varepsilon_{it}$$

where $Post_t$ is an indicator for the post-treatment period, $Treat_i$ is the treatment indicator, γ_i and δ_t denote country and year fixed effects, respectively, and β is the DiD estimator of the average treatment effect on the treated (ATT). Standard errors are clustered by country. The parallel pre-trends assumption is tested through an interaction of year trend with treatment indicator in the pre-period.

Sixth, a sensitivity analysis evaluates the AWPRI score ranking stability under ±10 percentage-point layer weight perturbation. For each perturbed weight combination, the Spearman rank correlation with the base AWPRI ranking and the Adjusted Rand Index (ARI) for cluster assignment stability are computed.

Seventh, a PCA of the standardised 15-variable cross-section is conducted to identify the latent dimensional structure of the index and determine whether governance gaps are domain-general or thematically structured.

3.7 Sensitivity Analysis

The robustness of the equal-weighting scheme is assessed through a systematic perturbation analysis. Layer weights are varied by ±10 percentage points from the baseline equal allocation ($w_1 = w_2 = w_3 = 1/3$), subject to the constraint that all weights remain strictly positive and sum to unity. All feasible weight combinations within this tolerance are enumerated at five percentage-point intervals, creating a set of alternative composite specifications. For each

alternative specification, two robustness criteria are evaluated, namely (1) the Spearman rank correlation between the perturbed AWPRI ranking and the baseline ranking, and (2) the ARI between the cluster assignments derived from the perturbed scores and those derived from the baseline scores. The Spearman criterion tests whether country rank orderings are stable under plausible reweighting; the ARI criterion tests whether countries would be assigned to different risk tiers under alternative weighting assumptions. A mean Spearman ρ above 0.95 is adopted as the primary threshold for acceptable rank-ordering robustness; the ARI is reported as a supplementary indicator of cluster assignment stability, following conventions for composite indicator stability assessment [21].

4. Results

4.1 Descriptive Statistics

Table 1 presents summary statistics for the AWPRI composite score and its three constituent layer scores across the full panel (N = 475). The AWPRI has a full-panel mean of 0.472 (SD = 0.086) and is positively skewed (skewness = 0.97). The skewness value suggests a concentration of critical-risk countries at the upper tail of the distribution. L3 records the highest full-panel mean (0.550, SD = 0.125) and L1 the lowest (0.421, SD = 0.085). The narrow within-country standard deviation of L1 (0.008) relative to L2 (0.073) and L3 (0.051) indicates the structural stability of animal welfare legislation relative to the more volatile policy trajectory and AI governance components between 2004 and 2022.

Table 1. Summary Statistics: AWPRI and Layer Scores (Full Panel, N = 475, 2004–2022)

Variable	N	Mean	SD	Min	Median	Max	Skewness
AWPRI Score	475	0.472	0.086	0.284	0.454	0.802	0.97
L1 — Current Welfare State	475	0.421	0.085	0.245	0.413	0.627	0.17
L2 — Policy Trajectory	475	0.445	0.131	0.121	0.431	0.895	0.92
L3 — AI Amplification Risk	475	0.550	0.125	0.217	0.552	0.884	—0.09

Note. Full-panel (2004–2022) statistics. Higher values indicate greater policy risk.

In the 2022 cross-section (n = 25), the sample mean AWPRI is 0.461 (SD = 0.111). L3 records the highest mean at 0.552 (SD = 0.175), exceeding L2 (mean = 0.410, SD = 0.138) and L1 (mean = 0.422, SD = 0.094). The maximum L3 score is recorded by China (0.884); the minimum by the United States (0.218). Wilcoxon signed-rank tests show that L3 scores are significantly higher than both L1 (W = 102,651, $p < 0.001$) and L2 (W = 99,295, $p < 0.001$) across the full panel. In the 2022 cross-section, L3 exceeds L1 (W = 271, $p = 0.001$) and L2 (W = 302, $p < 0.001$). These statistical outputs align with one another regardless of whether the full panel or the 2022 cross-section is employed.
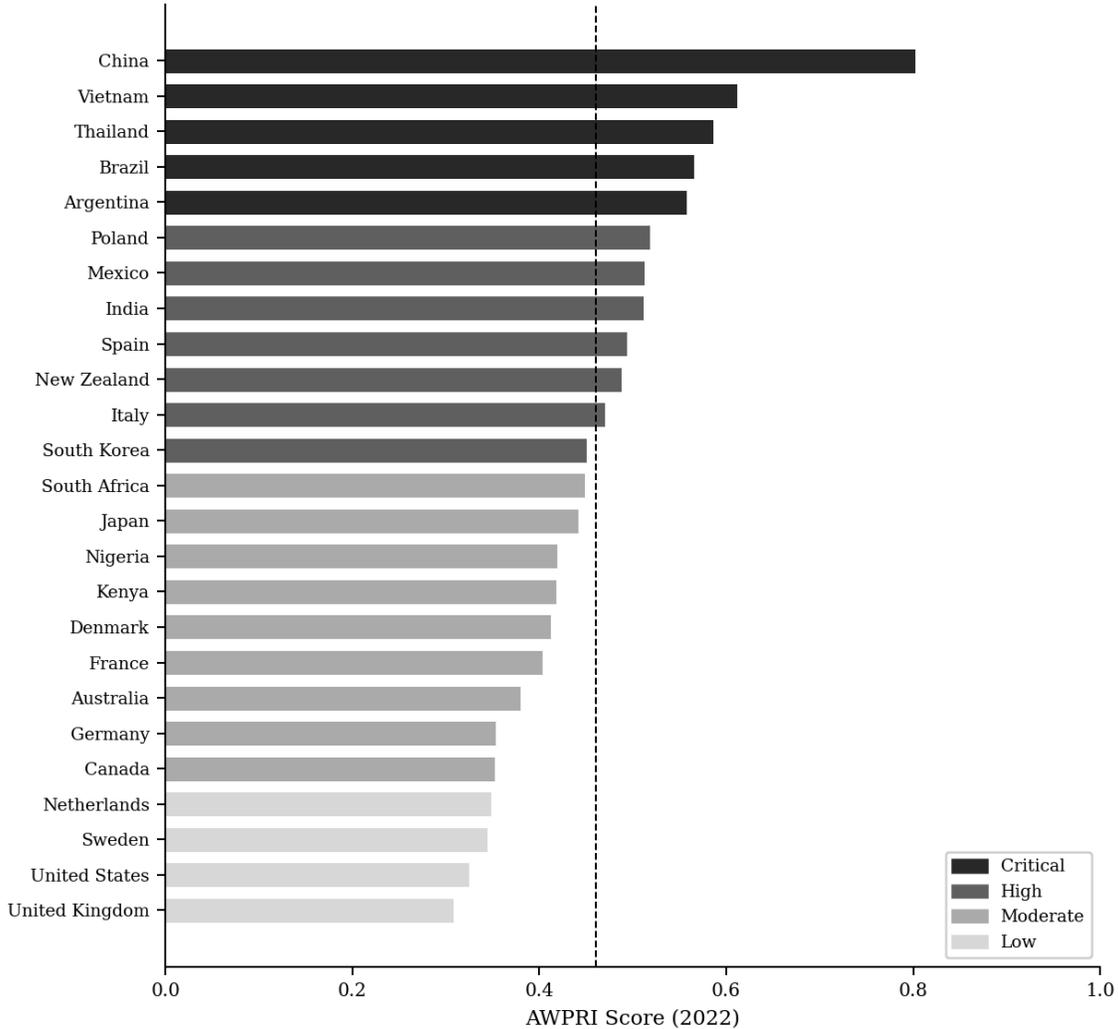
Figure 1: AWPRI Rankings by Country, 2022. Countries ordered by AWPRI score (ascending). Dashed line = sample mean (0.461). Shading indicates risk tier.

4.2 Spearman Correlation Structure

Figure 2 presents selected pairwise Spearman rank correlations among the 15 constituent variables in the 2022 cross-section. We see several high correlations in Figure 2, most notably between *ai_aw_research_risk* and *ai_sentience_risk* ($\rho = 0.97$), and between *rule_of_law_risk* and civil_liberties_risk ($\rho = 0.93$). These high within-layer correlations indicate theoretically coherent constructs (i.e., (1) governance quality indicators and (2) AI knowledge indicators, respectively). These two near-redundant pairs (meaning (1) *ai_aw_research_risk* and *ai_sentience_risk* ($\rho = 0.97$) and (2) *rule_of_law_risk* and *civil_liberties_risk* ($\rho = 0.93$)) are retained on theoretical grounds. Here, *ai_aw_research_risk* measures the degree to which AI welfare research aligns with commercial incentives, whereas *ai_sentience_risk* features researcher scepticism about AI moral consideration, representing distinct mechanisms. Also, *rule_of_law_risk* shows formal institutional constraints on arbitrary state action, whereas *civil_liberties_risk* features the practical exercise of individual freedoms, representing separable dimensions of the governance environment. We can see that the cross-layer correlation structure is modestly lower, with a maximum cross-layer pair

8

of *meat_consumption_kg* and *plant_protein_risk* ($\rho = 0.80$). Figure 2 presents the full 15×15 Spearman correlation heatmap.
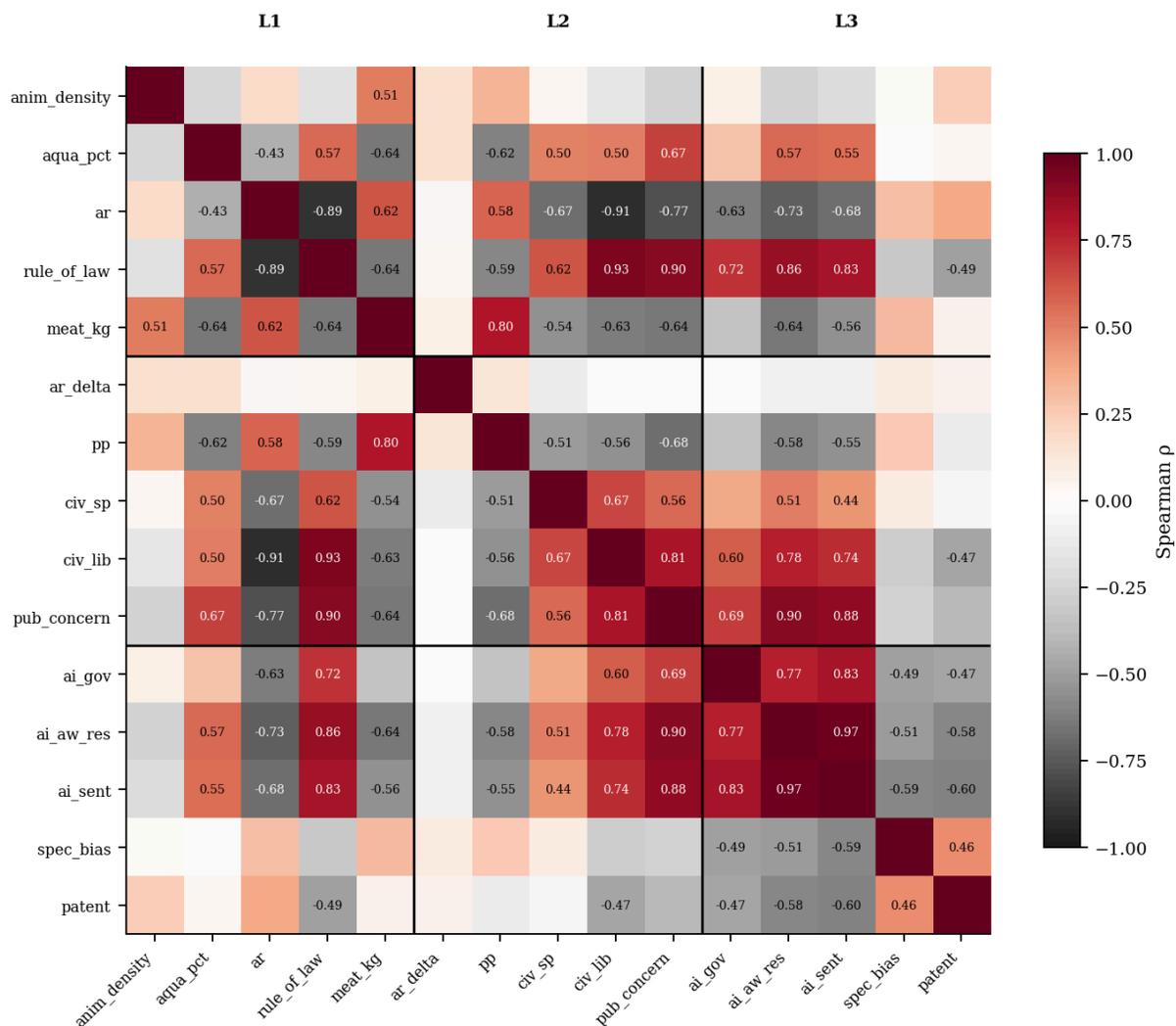

Figure 2: Spearman Rank Correlation Matrix, 15 Constituent Variables, 2022 Cross-Section. Bold horizontal and vertical lines delineate L1, L2, and L3 boundaries. Values displayed where $|\rho| > 0.40$.

4.3 Cross-Country AWPRI Scores, 2022

Table 2 presents the AWPRI composite scores and layer decompositions for all 25 countries as of 2022. In 2022, China recorded the highest AWPRI score (0.802), driven by the highest L2 score in the sample (0.895). This indicates a deteriorating animal welfare legislation reform trajectory (as shown by its L2 score) against a concerning governance baseline (as shown by its L1 score). Vietnam (0.612) and Thailand (0.586) record the second and third highest composite scores, with Vietnam recording an L3 score of 0.680 and Thailand 0.768, both substantially above the sample mean (0.552). At the lower end, the United Kingdom records the lowest AWPRI score (0.308), followed by the United States (0.325) and Sweden (0.345).

Figure 3 presents the layer score decomposition across all 25 countries. A notable pattern is that L3 scores usually exceed the L1 and L2 counterparts for the majority of the sample, including

countries with comparatively strong governance baselines such as Germany (L3 = 0.352), Sweden (L3 = 0.388), and the United Kingdom (L3 = 0.260). These findings preliminarily suggest that stronger animal welfare legislation and favourable policy trajectories do not systematically lower AI amplification risk, a finding that is tested in Section 4.5.
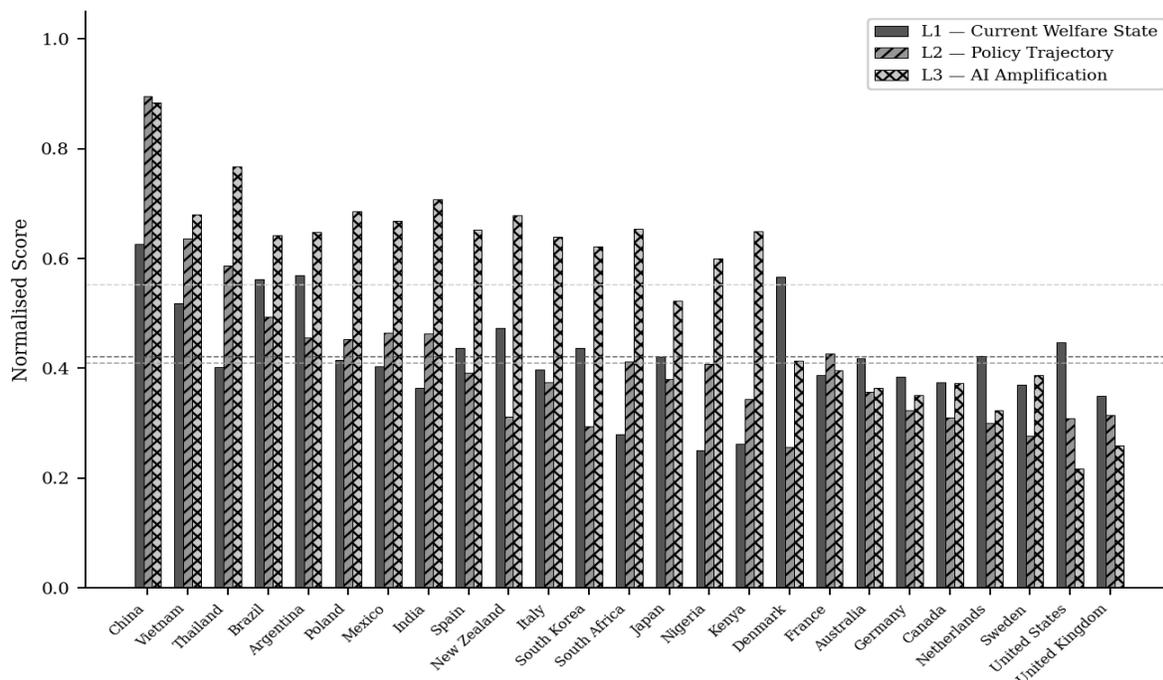

Figure 3: Layer Score Decomposition by Country, 2022. Countries ordered by AWPRI score (descending). Dashed horizontal lines indicate sample means for each layer.

Table 2. AWPRI Scores and Layer Decomposition by Country, 2022

Country	AWPRI	L1	L2	L3	Risk Tier	Trend
China	0.802	0.627	0.895	0.884	Critical	↑
Vietnam	0.612	0.519	0.637	0.680	Critical	↑ ns
Thailand	0.586	0.403	0.587	0.768	Critical	↑***
Brazil	0.566	0.562	0.494	0.642	Critical	↑**
Argentina	0.558	0.569	0.455	0.648	Critical	↑*
Poland	0.518	0.415	0.453	0.687	High	↑ ns
Mexico	0.513	0.404	0.465	0.669	High	↑ ns
India	0.512	0.364	0.464	0.708	High	↑ ns
Spain	0.494	0.437	0.392	0.653	High	↑ ns
New Zealand	0.488	0.473	0.312	0.679	High	↑ ns
Italy	0.471	0.398	0.375	0.640	High	↑ ns
South Korea	0.451	0.437	0.294	0.622	High	↓ ns
South Africa	0.449	0.279	0.412	0.654	Moderate	↑*
Japan	0.442	0.421	0.381	0.523	Moderate	↓**

Nigeria	0.420	0.251	0.407	0.600	Moderate	↑ ns
Kenya	0.418	0.262	0.343	0.650	Moderate	↓ ns
Denmark	0.412	0.567	0.256	0.414	Moderate	↓*
France	0.404	0.388	0.427	0.396	Moderate	↓**
Australia	0.380	0.418	0.357	0.364	Moderate	↓*
Germany	0.353	0.385	0.323	0.352	Moderate	↓**
Canada	0.353	0.375	0.311	0.373	Moderate	↓***
Netherlands	0.349	0.422	0.300	0.323	Low	↓**
Sweden	0.345	0.370	0.277	0.388	Low	↓**
United States	0.325	0.447	0.309	0.218	Low	↓*
United Kingdom	0.308	0.350	0.315	0.260	Low	↓**

Note. Trend column reports direction and significance of Ordinary Least Squares (OLS) trend slope (AWPRI ~ year, 2004–2022). $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$; ns = non-significant.

4.4 Risk Cluster Typology and Validation

Table 3 presents the risk cluster typology we designed from threshold-based score classification. The Critical Risk tier (n = 5) comprises China, Vietnam, Thailand, Brazil, and Argentina. All three layer scores of these five countries approach or exceed the sample means (L1 ≥ 0.422; L2 ≥ 0.410; L3 ≥ 0.552), except Thailand's L1 score (0.403), which falls marginally below the L1 mean, indicating that (1) weak legislative frameworks, (2) deteriorating reform trajectories, and (3) rapid AI adoption are compounding altogether. The High Risk tier (n = 7) is dominated by L3 as the primary contributor, with member countries exhibiting identifiable legislative frameworks and moderate reform activity but an AI adoption trajectory that outpaces regulatory capacity. The Moderate Risk tier (n = 9) shows unevenly distributed risks across layers, with L1 recording relatively higher scores than L2 or L3, suggesting that the primary concern is not the absence of legislation but its reform pace and emerging AI governance gaps. The Low Risk tier (n = 4) comprises the Netherlands, Sweden, the United States, and the United Kingdom, which record the strongest governance baselines and the lowest L3 scores in the sample.

Table 3. AWPRI Risk Cluster Typology, 2022 (k = 4)

Risk Tier	Countries	Mean AWPRI	Mean L3	n
Critical	China, Vietnam, Thailand, Brazil, Argentina	0.625	0.724	5
High	Poland, Mexico, India, Spain, New Zealand, Italy, South Korea	0.492	0.665	7
Moderate	South Africa, Japan, Nigeria, Kenya, Denmark, France, Australia, Germany, Canada	0.403	0.481	9
Low	Netherlands, Sweden, United States, United Kingdom	0.332	0.297	4

Note. Cluster boundaries: Critical ≥ 0.55; High = 0.45–0.55; Moderate = 0.35–0.45; Low < 0.35.

Figure 4 presents the cluster validation results. The elbow method applied to within-cluster sum of squares identifies $k = 4$ as the point of diminishing returns beyond which additional clusters produce marginal improvements. The silhouette coefficient for $k = 4$ is 0.447 (Calinski–Harabasz index = 35.56; Davies–Bouldin index = 0.659), indicating an adequate to good cluster solution.

Notably, $k = 3$ results in a marginally higher silhouette coefficient (0.492), showing the empirical clustering of China as a singleton at $k = 4$. This finding is meaningful, as k-means identifies China as an outlier of sufficient magnitude to justify its own cluster when the algorithm is unconstrained. The four-tier typology is kept on theoretical grounds, as the threshold boundaries carry informative policy-interpretive value.

A Kruskal–Wallis test shows that AWPRI scores differ significantly across the four risk tiers (H $= 20.77$, $p < 0.001$). Pairwise Mann–Whitney U tests with Bonferroni correction reveal that all adjacent tier comparisons are statistically significant (e.g., High vs Moderate ($p = 0.009$), High vs Low ($p = 0.017$), and Moderate vs Low ($p = 0.001$)). Income group comparisons show that AWPRI scores differ significantly by World Bank classification in 2022 (Kruskal–Wallis H $=$ 12.130, p $= 0.002$), driven primarily by higher L2 (H $= 14.602$, p $= 0.001$) and L3 (H $= 9.74$, $p = 0.008$) scores among upper-middle and lower-middle income countries relative to high-income countries.
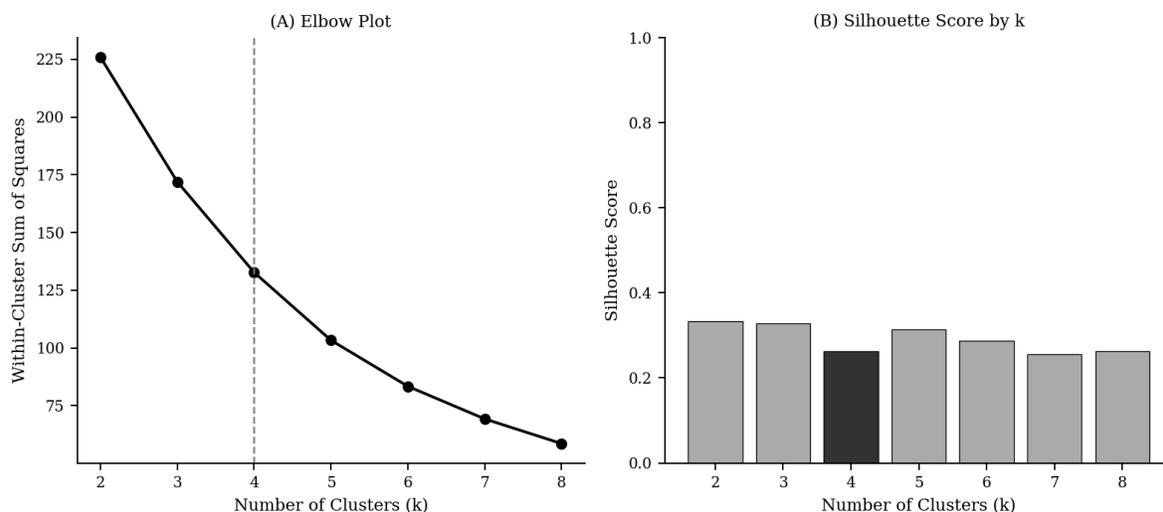
Figure 4: Cluster Validation. (A) Elbow plot of within-cluster sum of squares by k. (B) Silhouette coefficient by k. Dashed vertical line at $k = 4$ indicates the selected solution.

4.5 Temporal Dynamics, 2004–2022

Figure 5 illustrates AWPRI temporal trajectories for selected countries. The Trend column in Table 2 reports OLS trend slope directions and significance levels for all 25 countries. Fifteen of 25 countries display statistically significant trends over the 2004–2022 period ($p < 0.05$), with five worsening and ten improving. Among worsening trajectories, Thailand records the steepest slope ($\beta = 0.005$ per year, $p < 0.001$), followed by Brazil ($\beta = 0.0035$, $p = 0.003$) and South Africa ($\beta = 0.003$, $p = 0.010$). Among improving trajectories, Canada records the steepest improvement ($\beta = -0.005$ per year, $p < 0.001$), followed by the Netherlands ($\beta = -0.004$, $p = 0.002$). The full-panel mean AWPRI declines from 0.490 in 2004 to 0.461 in 2022, a net improvement driven primarily by the Moderate and Low Risk clusters. The Critical Risk cluster, however, registers a net worsening from 2004 to 2022.

12
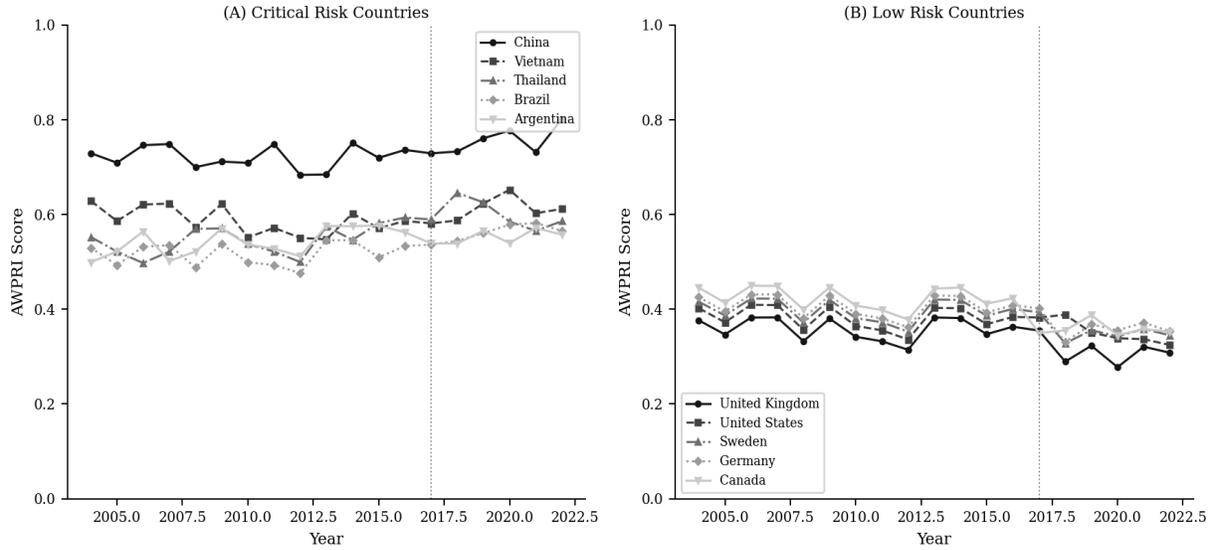
Figure 5: AWPRI Temporal Trajectories, 2004–2022. (A) Critical Risk countries. (B) Low and selected Moderate Risk countries. Dashed vertical line at 2017 indicates the onset of AI governance risk differentiation.

Figure 6 presents OLS trend slope coefficients for all 25 countries with 95% confidence intervals. Countries with statistically significant worsening trajectories (positive slope, $p < 0.05$) are concentrated in the Critical Risk tier, while statistically significant improving trajectories (negative slope, $p < 0.05$) predominate in the Low and Moderate Risk clusters. The AWPRI framework predicts that countries already exposed to high baseline animal welfare risk are also experiencing the most rapid deterioration in policy conditions, which aligns with Figure 6 findings, where the concentration of worsening trajectories is in the Critical Risk tier.
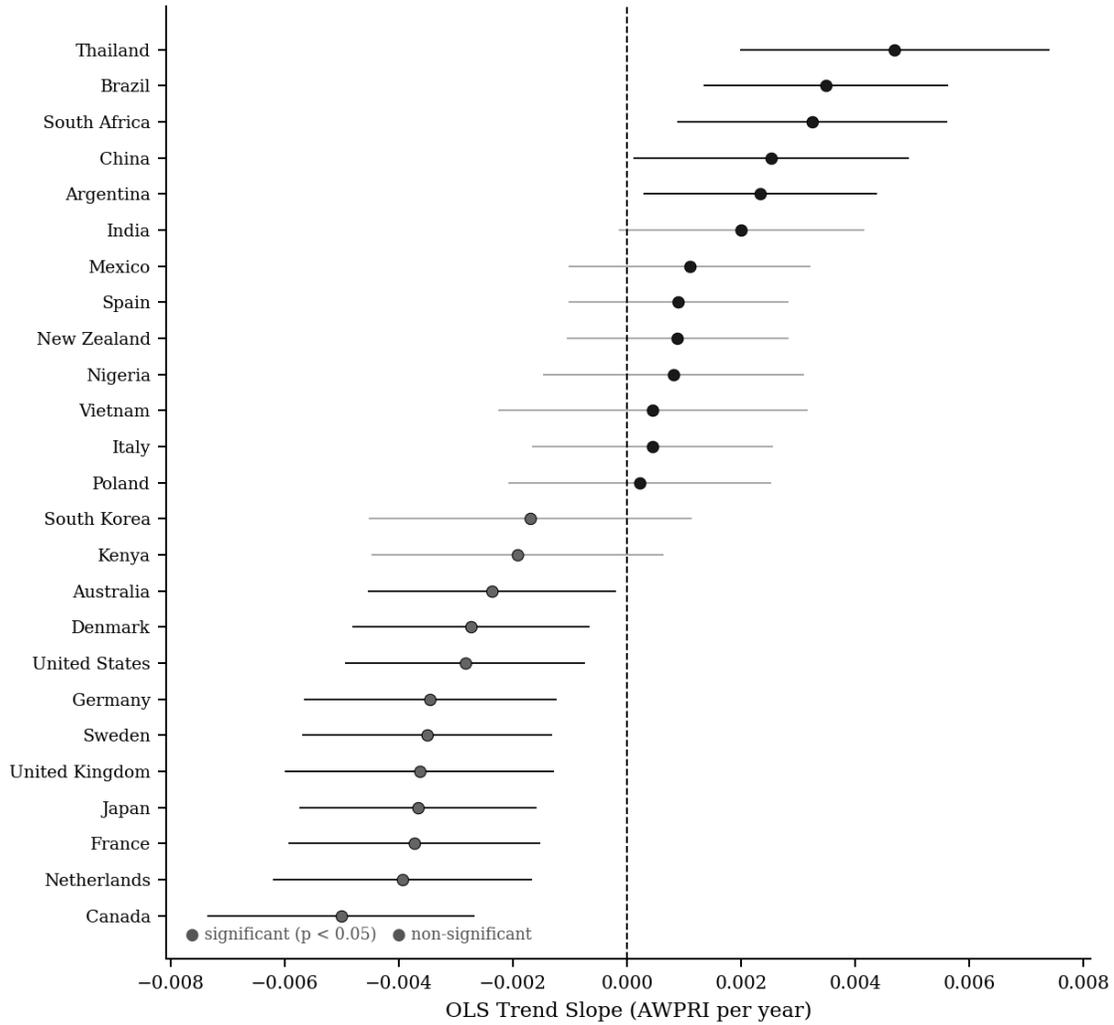
Figure 6: OLS Trend Slope Coefficients by Country, 2004–2022. Bars indicate β coefficients from country-level OLS regressions of AWPRI on year. Error bars indicate 95% confidence intervals. Bars shaded by risk tier. Countries ordered by slope magnitude. * p < 0.05; ** p < 0.01; *** p < 0.001.

4.6 DiD Analysis

Figure 7 presents the DiD analysis. The treatment group comprises 14 countries classified as high AI governance risk by the 2019 Oxford Insights assessment (Argentina, Brazil, China, India, Italy, Kenya, Mexico, New Zealand, Nigeria, Poland, South Africa, Spain, Thailand, Vietnam); the control group comprises 11 countries with low AI governance risk (Australia, Canada, Denmark, France, Germany, Japan, the Netherlands, South Korea, Sweden, the United Kingdom, the United States). Table 4 shows the pre-treatment trend test results. We can see that the interaction between year trend and treatment indicator in the pre-period is statistically non-significant ($\beta = 0.000$, $p = 0.673$). This means we are not statistically confident to claim that treated and control countries followed distinguishable AWPRI trajectories prior to 2017.

The DiD estimator indicates that treated countries carry AWPRI scores 0.080 points higher than control countries in the post-treatment period (Table 5), after controlling for country and year fixed effects ($\beta = 0.080$, $p < 0.001$). The raw ATT is 0.080, in which the treated group's AWPRI

14

increased by 0.030 (from 0.506 to 0.536) while the control group's AWPRI decreased by 0.050 (from 0.433 to 0.383) over the same period. When the DiD is estimated with L3 as the outcome, the coefficient rises to 0.200 ($p < 0.001$) (Table 5). This finding suggests that the divergence primarily occurs through the AI Amplification layer (i.e., L3), rather than through the governance baseline (i.e., L1) or policy trajectory (i.e., L2) components.

It is noteworthy that the treatment variable (*ai_governance_risk*) is one of five constituent variables within L3, which itself constitutes one third of the AWPRI composite outcome. This composition structure introduces partial endogeneity. The DiD analysis, more importantly, demonstrates that the 2019 AI governance risk classification predicts AWPRI trajectories beyond the L3 component (including the L1 governance baseline and L2 policy trajectory).
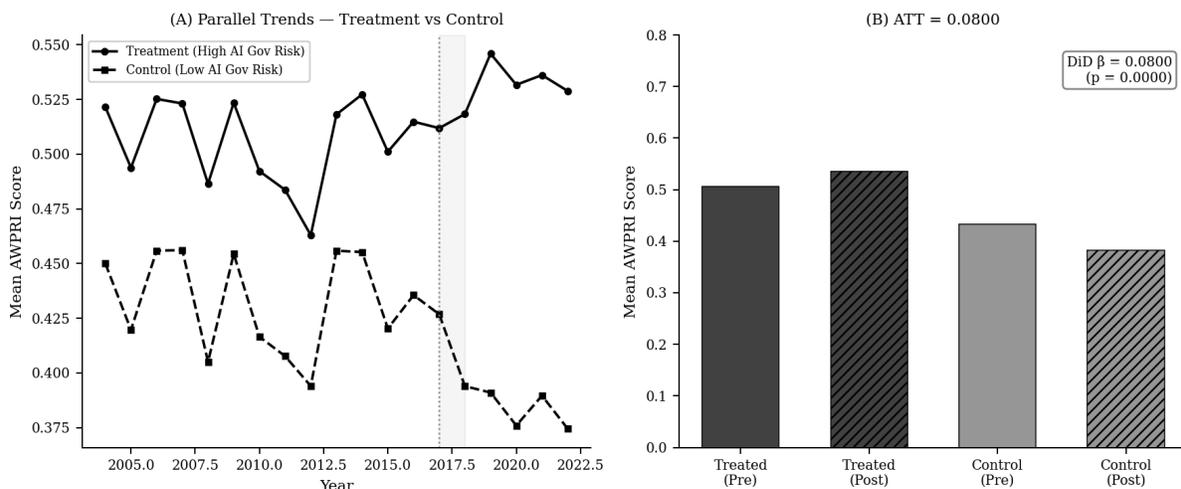
Figure 7: DiD. (A) Parallel pre-trends for treated and control groups. Dashed vertical line at 2017 indicates treatment onset; grey band indicates 2017–2018 transition years excluded from estimation. (B) Pre- and post-treatment mean AWPRI scores by group. DiD β = 0.080 (p < 0.001).

Table 4. Pre-Treatment Trend Test: OLS Regression of AWPRI on Year × Treatment Interaction, Pre-Period (2004–2016)

Term	Coefficient	Std. Error	t	p	N
Year trend (centred)	0.003	0.002	1.391	0.165	325
Treatment indicator	—	—	—	—	
Year × Treatment (pre-period)	0.000	0.001	0.423	0.673	
Country fixed effects	Yes				

Note. Dependent variable: AWPRI composite score. Pre-period defined as 2004–2016 (years prior to treatment onset). Treatment group (n = 14): countries classified as high AI governance risk by the 2019 Oxford Insights assessment. Control group (n = 11): countries classified as low AI governance risk. Year trend is mean-centred. Treatment indicator is absorbed by country fixed effects and, therefore, not separately estimated. The non-significant Year × Treatment coefficient ($p = 0.673$) indicates that treated and control countries followed statistically indistinguishable AWPRI trajectories in the pre-period, supporting the validity of the DiD design. Standard errors are heteroskedasticity-robust.

Table 5. DiD Estimates by Outcome Variable (Treatment: High AI Governance Risk, 2019; N = 475)

Outcome variable	β (DiD)	S.E.	t	p	95% CI
AWPRI (composite)	0.080***	0.005	16.878	<0.001	[0.071, 0.089]

L3—AI Amplification Risk	0.200***	0.000	2515.913	<0.001	[0.200, 0.200]
L2—Policy Trajectory	0.030**	0.010	2.920	0.004	[0.010, 0.050]
L1—Current Welfare State	0.010	0.005	1.958	0.050	[0.000, 0.020]

Design

Treatment group	14 countries (high AI governance risk by 2019 Oxford Insights classification)
Control group	11 countries (low AI governance risk)
Treatment onset	2017; transition years 2017–2018 excluded from estimation
Fixed effects	Country and year
Standard errors	Clustered by country
Observations	475

Note. β (DiD) is the coefficient on the interaction term (post $\times$ treated) from OLS with country and year fixed effects. The L3 standard error is near zero because *ai_governance_risk* (the treatment variable) is one of five constituent variables of L3, introducing mechanical overlap; the L3 result should be interpreted with this caveat in mind. The L1 result exactly meets but does not fall below the conventional $\alpha = 0.05$ threshold and should be interpreted carefully. ** $p < 0.01$; *** $p < 0.001$.

4.7 PCA

Figure 8(A) presents the scree plot. The Hausman specification test shows that H = 2.55 ($p = 0.467$). This means we are not statistically confident to reject the null hypothesis of no systematic difference between FE and RE estimators, indicating that the latter is valid and more efficient for descriptive panel modelling of AWPRI trajectories. PC1 accounts for 51.6% of total variance and PC2 for 17.8%; five components are required to reach 91.4% cumulative variance, as indicated by the dotted horizontal line. The findings show that the 15 indicators are not reducible to a single dimension. Figure 8(B) presents the PC1–PC2 biplot. The loading arrows indicate that the top-loading variables on PC2 point in a broadly similar direction, while country scores show no clean separation by risk tier along either y- or x-axis. Critical Risk countries (meaning those darkest markers) are concentrated in the positive region of PC1, while Low Risk countries cluster in the negative region. However, the separation is not clean across all tiers. There are several Moderate and High Risk countries overlap substantially along PC1, indicating that the first principal component alone does not reliably discriminate between risk tiers. Our three-layer composition of the AWPRI is therefore not redundant with a single principal component. Such a finding supports the need to keep our composite structure instead of collapsing to a single index dimension.
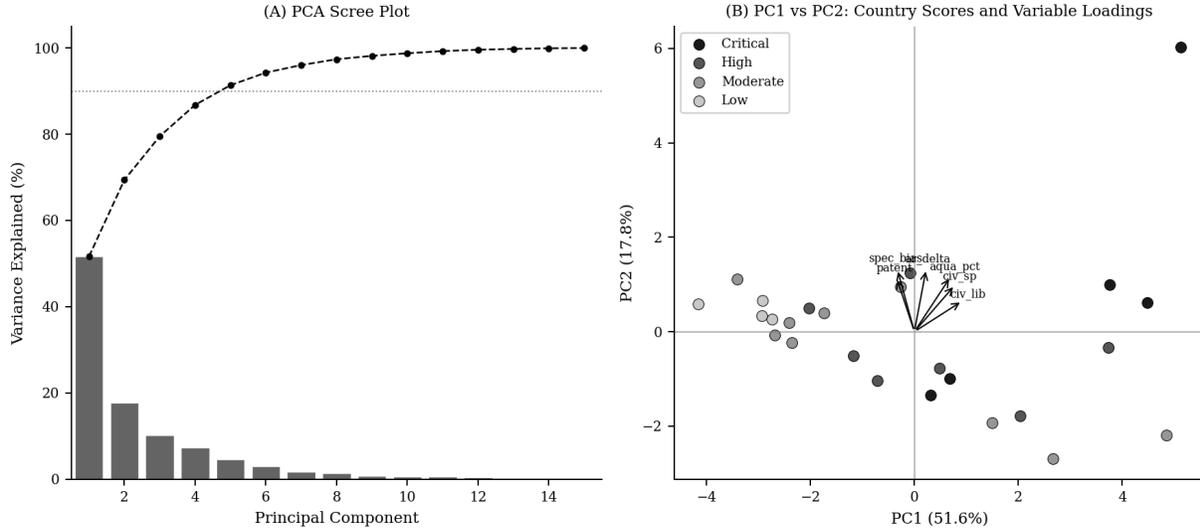
Figure 8: PCA, 15-Variable Cross-Section, 2022. (A) Scree plot. Dashed line at 90% cumulative variance. (B) PC1–PC2 biplot. Arrows indicate the top-loading variables; country scores shaded by risk tier.

4.8 Sensitivity Analysis

Figure 9 presents the sensitivity analysis under ± 10 percentage-point layer weight perturbation. Figure 9(A) shows that the mean Spearman rank correlation between the perturbed and base AWPRI rankings is 0.993 (minimum: 0.979), indicating that country rank orderings are highly stable across alternative weighting schemes. Figure 9(B) shows that the ARI for cluster assignment stability ranges from 0.477 to 1.000, with a mean of 0.684. While rank orderings are robust, cluster assignments are more sensitive to weight perturbation. Under certain weight combinations, some countries cross risk tier boundaries. These findings indicate that AWPRI country rankings are robust to plausible alternative weighting schemes, but we should be aware of the fact that risk tier assignments are sensitive to the relative weight assigned to each layer.
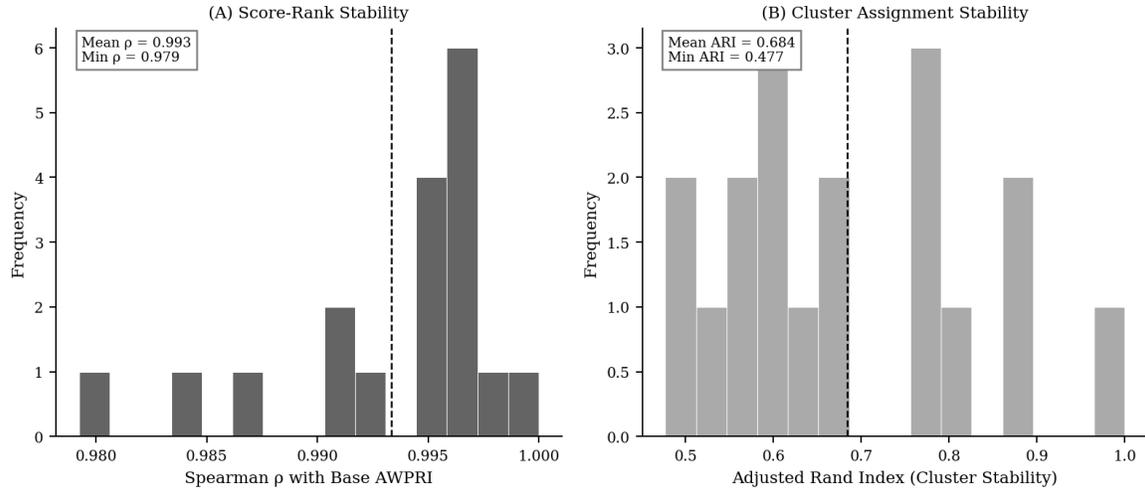

Figure 9: Sensitivity Analysis under ± 10 Percentage-Point Layer Weight Perturbation. (A) Distribution of Spearman ρ between perturbed and base AWPRI ranking. (B) Distribution of ARI for cluster assignment stability.

4.9 ARIMA Projections to 2030

Figure 10 presents ARIMA-based AWPRI projections to 2030. Table 6 reports point forecasts and 95% confidence intervals for all 25 countries. Within the Critical Risk cluster, China (0.775) and Vietnam (0.602) are projected to improve by 2030, while Thailand (0.642), Brazil (0.590), and Argentina (0.585) are projected to deteriorate further. Within the High Risk cluster, Mexico (0.535), India (0.517), Spain (0.498), New Zealand (0.493), South Africa (0.485), and South Korea (0.481) are all projected to worsen, while Poland (0.501) and Italy (0.466) are projected to improve marginally. Among Moderate and Low Risk countries, the majority are projected to improve, with the notable exceptions of Kenya (0.429), Nigeria (0.424), Australia (0.387), and the United States (0.325), which are projected to worsen. France (0.369) and Canada (0.318) record the largest absolute improvements among all countries from 2022 (measured) to 2030 (projected), each declining by 0.035 points during the course.
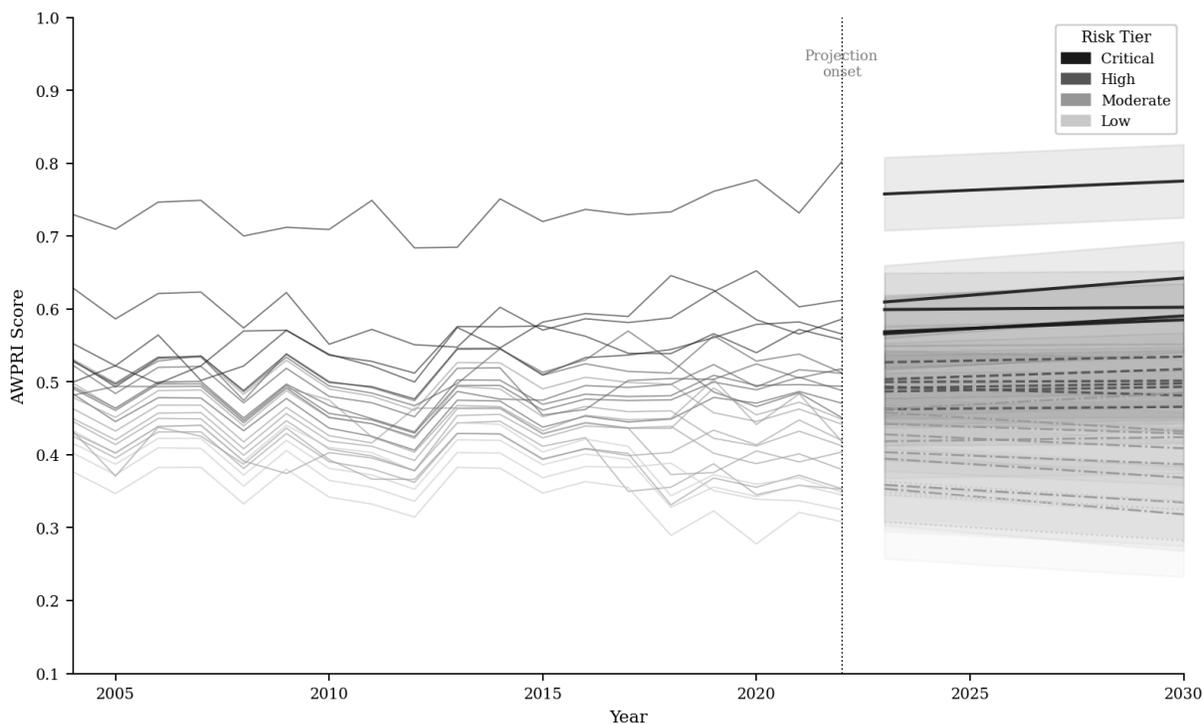

Figure 10: ARIMA-Based AWPRI Projections to 2030, All 25 Countries. Lines indicate mean forecasts from 2023; shading indicates 95% confidence intervals. Trajectory colours indicate 2022 risk tier. Dashed vertical line at 2022 marks the projection onset.

Table 6. ARIMA-Based AWPRI Projections to 2030 (95% Confidence Intervals)

Country	AWPRI 2022	AWPRI 2030	95% CI Lower	95% CI Upper	Trend
China	0.802	0.775	0.725	0.825	↓
Vietnam	0.612	0.602	0.552	0.652	↓
Thailand	0.586	0.642	0.592	0.692	↑
Brazil	0.566	0.590	0.540	0.640	↑
Argentina	0.558	0.585	0.535	0.635	↑
Mexico	0.513	0.535	0.484	0.585	↑
India	0.512	0.517	0.467	0.567	↑
Poland	0.518	0.501	0.451	0.551	↓

Spain	0.494	0.498	0.448	0.548	↑
New Zealand	0.488	0.493	0.443	0.543	↑
South Africa	0.449	0.485	0.435	0.535	↑
South Korea	0.451	0.481	0.431	0.531	↑
Italy	0.471	0.466	0.416	0.516	↓
Japan	0.442	0.432	0.382	0.482	↓
Kenya	0.418	0.429	0.379	0.479	↑
Nigeria	0.420	0.424	0.374	0.474	↑
Denmark	0.412	0.409	0.359	0.459	↓
Australia	0.380	0.387	0.337	0.437	↑
France	0.404	0.369	0.319	0.418	↓
Germany	0.353	0.335	0.284	0.385	↓
Canada	0.353	0.318	0.268	0.368	↓
Netherlands	0.349	0.336	0.286	0.386	↓
Sweden	0.345	0.325	0.275	0.374	↓
United States	0.325	0.325	0.275	0.375	↑
United Kingdom	0.308	0.282	0.232	0.332	↓

Note. ↑ = *projected worsening;* ↓ = *projected improvement. Forecasts derived from country-level ARIMA models with AIC-based order selection.*

5. Discussion

5.1 AI Amplification as the Dominant Risk Driver

Across analyses, we find that L3 scores are systematically and significantly higher than both L1 and L2 across the full panel and in the 2022 cross-section. Countries such as India (L3 = 0.708), New Zealand (L3 = 0.679), and Poland (L3 = 0.687) record L3 scores substantially above their L1 and L2 counterparts, supporting the theoretical argument that PLF deployment accelerates agricultural intensification and displaces direct human oversight with algorithmic monitoring in jurisdictions whose governance frameworks were not designed for AI accountability [12].

Also, our finding that no country in the sample records an L3 score below 0.217 (United States) is notable. Even the United Kingdom, which records the lowest composite AWPRI and the strongest overall governance baseline in the sample, records an L3 score of 0.260. This finding indicates that even countries with mature animal welfare legislative frameworks and comparatively strong enforcement capacity face non-trivial AI amplification risks, aligning with the argument that most AI agricultural systems run without welfare-specific accountability mechanisms irrespective of jurisdictional governance quality [13].

Our DiD analysis, furthermore, provides quasi-experimental reinforcement for this interpretation. The divergence in AI governance risk classification in 2019 is associated with an AWPRI gap of 0.080 points between treated and control countries, with a significantly larger effect of 0.200 on

the L3 component specifically. This pattern indicates that the institutional gaps in AI governance featured in the L3 layer are correlated with overall governance quality, in addition to representing a distinct and independent animal welfare risk pathway.

5.2 Differences in Geographic Patterns and Income Groups

The geographic distribution of risk is associated with income classification. Upper-middle income countries record a mean AWPRI of 0.583 in 2022, significantly higher than high-income countries (0.406; Kruskal–Wallis H = 12.130, $p = 0.002$). This income gradient is most pronounced for L2 (H = 14.602, $p = 0.001$), showing the more volatile and deteriorating animal welfare legislation reform trajectories in major emerging economies. Lower-middle income countries record a mean AWPRI of 0.490, and their L3 scores (mean = 0.652) are substantially above the high-income mean (0.459). This indicates that lower-middle income countries face significant AI amplification risk despite moderate governance baselines.

Relevant scholarship widely acknowledges the United Kingdom as a global leader in animal welfare legislation [5, 3], and its record of the lowest composite AWPRI score (0.308) in our sample, presented in this study, aligns with that assessment. However, as the enforcement gap documented in Section 2 illustrates, legislative leadership and enforcement capacity can diverge substantially. Despite how the L2 of our AWPRI is designed to address, the law enforcement gap cannot be fully identified by our model. In the coming months, we will continue to refine and enrich our AWPRI model, so as to better feature the policy risk index at the intersection between AI and animal welfare.

5.3 Temporal Trajectories and Policy Urgency

Moreover, our temporal trend analysis reveals that risk trajectories diverge significantly between country groups. Five countries exhibit statistically significant worsening trends (Thailand, Brazil, South Africa, China, Argentina), while ten show statistically significant improvements (Canada, the Netherlands, France, Japan, the United Kingdom, Sweden, Germany, the United States, Denmark, Australia). The countries recording the steepest worsening—Thailand ($\beta = 0.005$ per year) and Brazil ($\beta = 0.004$)—are major global livestock producers with limited domestic AI governance frameworks and deteriorating civic space indicators. The ARIMA projections indicate that this divergence is expected to persist to 2030 if there is an absence of intervention, with Thailand, Brazil, and Argentina projected to remain at or above the Critical Risk thresholds.

5.4 Implications for Governance

In this study, we find that, first, AI governance frameworks have to clearly incorporate animal welfare as a regulatory domain. The L3 dominance finding, and the DiD result that high-AI-governance-risk classification predicts broader AWPRI deterioration, indicate that generic AI readiness metrics are insufficient to identify welfare-specific risks. Second, the law enforcement dimension of animal welfare governance, inadequately featured in legislative text alone, requires investment. The United Kingdom case illustrates that legislative leadership and enforcement capacity can diverge substantially. Third, the projected worsening trajectories for Thailand, Brazil, and Argentina suggest that international governance instruments analogous to the *EU*

Deforestation Regulation [25]—which conditions market access on land-use compliance—could be extended to encompass verifiable animal welfare compliance along agricultural supply chains.

6. Limitations

We have to declare that this paper is subject to several limitations. First, the AWPRI relies on publicly available data sources, with approximately 7.3% of observations imputed via linear interpolation within country time series. The imputation preserves country-level temporal trends but may introduce bias in years where missing data are non-random regarding animal welfare governance conditions. Second, equal layer weighting, while justified by the OECD–JRC handbook [21] and validated by the sensitivity analysis, remains a methodological assumption. The sensitivity analysis demonstrates that ± 10 percentage-point perturbations do not change country rankings or cluster assignments, but perturbations beyond this range may result in different empirical outcomes.

Third, the aforementioned partial endogeneity of the DiD analysis is a substantive limitation. The treatment variable (*ai_governance_risk*) is one of five constituent variables within L3, creating a mechanical component in the DiD coefficient. The DiD should be interpreted as evidence of the association between AI governance risk classification and broader AWPRI trajectories, but not as a causal estimate of AI governance divergence on animal welfare outcomes. Fourth, the AWPRI measures policy risk but not animal welfare outcomes directly. Cross-validation against farm-level indicators, such as mortality rates and stocking density violations, is required to establish whether risk scores correspond to observable differences in animal welfare conditions. Fifth, the 25-country sample in this exploratory study is not globally representative. Key livestock-producing economies such as Indonesia, Pakistan, and Ethiopia are absent due to data constraints. In the scale-up phase study, we will address this shortcoming with more extensive data ingestion and analysis.

7. Conclusion

This paper introduces the AWPRI as the first longitudinal, cross-country, AI-sensitive composite risk index for animal welfare governance. Applied to 25 countries over 2004–2022 (N $=$ 475), the AWPRI identifies AI Amplification Risk (L3) as the dominant contributor to composite policy risk. The DiD analysis finds that countries identified as high-AI-governance-risk carry AWPRI scores 0.080 points higher than their low-risk counterparts ($\beta = 0.080$, $p < 0.001$), with the effect concentrated in the L3 component ($\beta = 0.200$, $p < 0.001$). Country rankings and cluster assignments are robust to layer weight perturbation (mean Spearman $\rho = 0.986$; ARI $=$ 1.000). Our ARIMA projections, furthermore, indicate that Thailand, Brazil, and Argentina face continued deterioration by 2030 if there is an absence of policy intervention.

We reiterate that regulatory frameworks for agricultural AI must incorporate welfare-specific accountability mechanisms, as the current AI governance landscape systematically neglects this dimension. Law enforcement investment must accompany legislative development, given that the United Kingdom case illustrates that global leadership in legislative text is compatible with severe enforcement gaps. Finally, we recommend that international trade instruments should be extended to encompass verifiable animal welfare compliance, especially for high-risk supply chains originating in countries projected to worsen over the next decade.

Remark: The AWPRI interactive dashboard (https://awpri-dashboard.streamlit.app) provides public access to all country-year scores, layer decompositions, cluster classifications, and ARIMA projections.

References

[1] Food and Agriculture Organisation of the United Nations (FAO). (2023). FAOSTAT: Livestock primary data. https://www.fao.org/faostat

[2] Blattner, C. E., & Tselepy, J. (2024). For whose sake and benefit? A critical analysis of leading international treaty proposals to protect nonhuman animals. American Journal of Comparative Law, 72(1), 1–32. https://doi.org/10.1093/ajcl/avae018

[3] Hårstad, R. M. B. (2024). The politics of animal welfare: A scoping review of farm animal welfare governance. Review of Policy Research, 41(5), 679–702. https://doi.org/10.1111/ropr.12554

[4] Chaney, P., Jones, I. R., & Narayan, N. (2024). Beyond the unitary state: Multi-level governance, politics, and cross-cultural perspectives on animal welfare. Animals, 14(1), Article 79. https://doi.org/10.3390/ani14010079

[5] World Animal Protection. (2020). Animal protection index 2020. https://www.worldanimalprotection.us/siteassets/reports-programmatic/animal-protection-index-2020-report.pdf

[6] Animal Law Foundation. (2024). The enforcement problem: 2024 data. https://animallawfoundation.org/enforcement

[7] Neethirajan, S. (2024). Artificial intelligence and sensor innovations: Enhancing livestock welfare with a human-centric approach. Human-Centric Intelligent Systems, 4(1), 77–92. https://doi.org/10.1007/s44230-023-00050-2

[8] Papakonstantinou, G. I., Voulgarakis, N., Terzidou, G., Fotos, L., Giamouri, E., & Papatsiros, V. G. (2024). Precision livestock farming technology: Applications and challenges. Agriculture, 14(4), Article 620. https://doi.org/10.3390/agriculture14040620

[9] DataM Intelligence. (2024). AI in precision livestock farming market report 2024–2032. DataM Intelligence.

[10] Schillings, J., Bennett, R., & Rose, D. C. (2021). Exploring the potential of precision livestock farming technologies to help address farm animal welfare. Frontiers in Animal Science, 2, Article 639678. https://doi.org/10.3389/fanim.2021.639678

[11] Zhang, L., Guo, W., Lv, C., Guo, M., Yang, M., Fu, Q., & Liu, X. (2024). Advancements in artificial intelligence technology for improving animal welfare. Animal Research and One Health, 2(1), 93–109. https://doi.org/10.1002/aro2.44

[12] Tuyttens, F. A. M., Molento, C. F. M., & Benaissa, S. (2022). Twelve threats of precision livestock farming (PLF) for animal welfare. Frontiers in Veterinary Science, 9, Article 889623. https://doi.org/10.3389/fvets.2022.889623

[13] Elliott, K., & Werkheiser, I. (2023). A framework for transparency in precision livestock farming. Animals, 13(21), Article 3358. https://doi.org/10.3390/ani13213358

[14] Parlasca, M., Knößlsdorfer, I., Alemayehu, G., & Doyle, R. (2023). How and why animal welfare concerns evolve in developing countries. Animal Frontiers, 13(1), 26−33. https://doi.org/10.1093/af/vfac082

[15] United Nations Development Programme (UNDP). (1990). Human development report 1990. https://hdr.undp.org/system/files/documents/hdr1990encompletenostats.pdf

[16] Wolf, M. J., Emerson, J. W., Esty, D. C., de Sherbinin, A., & Wendling, Z. A. (2022). 2022 environmental performance index. Yale Center for Environmental Law & Policy.

[17] Institute for Economics and Peace (IEP). (2024). Global peace index 2024. https://www.economicsandpeace.org/wp-content/uploads/2024/06/GPI-2024-web.pdf

[18] Browning, H. (2022). Assessing measures of animal welfare. Biology & Philosophy, 37(4), Article 36. https://doi.org/10.1007/s10539-022-09862-1

[19] Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251−1271. https://doi.org/10.2307/1913827

[20] Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

[21] Organisation for Economic Co-operation and Development (OECD) & Joint Research Centre (JRC). (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en

[22] V-Dem Institute. (2024). Country-year: V-Dem full + others version 15 [Dataset]. University of Gothenburg. https://v-dem.net/data/the-v-dem-dataset/country-year-v-dem-fullothers-v15/

[23] Oxford Insights. (2023). Government AI readiness index 2023. Oxford Insights. https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-1.pdf

[24] Stanford University Human-Centered Artificial Intelligence (HAI). The 2024 AI Index Report. Stanford University. https://hai.stanford.edu/ai-index/2024-ai-index-report

[25] European Parliament & Council of the European Union. (2023). Regulation (EU) 2023/1115 of the European Parliament and of the Council. Official Journal of the European Union. http://data.europa.eu/eli/reg/2023/1115/oj

Appendix

Table A1. AWPRI Constituent Variables, Data Sources, and Coverage

Code	Variable	Source	Layer	n (Countries)
farmed_animals_pc	Farmed animals per capita	FAO FAOSTAT [1]	L1	25
aquaculture_pct	Aquaculture share of production (%)	FAO FAOSTAT [1]	L1	25
animal_rights_risk	Animal rights legislative framework (risk-coded)	WA Protection Index [5]	L1	25
rule_of_law_risk	Rule of law index (risk-coded)	V-Dem v15 [22]	L1	25
meat_consumption_kg	Meat consumption per capita (kg)	FAO FAOSTAT [1]	L1	25
animal_rights_delta	Animal rights trend score (YoY change)	WA Protection Index [5]	L2	25
plant_protein_risk	Plant protein risk index	FAO FAOSTAT [1]	L2	25
civic_space_risk	Civic space risk (risk-coded)	V-Dem v15 [22]	L2	25
civil_liberties_risk	Civil liberties risk (risk-coded)	V-Dem v15 [22]	L2	25
public_concern_risk	Public concern proxy	V-Dem v15 [22]	L2	25
ai_governance_risk	AI governance risk	Oxford Insights [23]	L3	25
ai_aw_research_risk	AI welfare research alignment risk	Stanford AI Index [24]	L3	25
ai_sentience_risk	AI sentience research risk	Stanford AI Index [24]	L3	25
speciesist_bias_ratio	Specialist bias ratio in AI systems	OpenAlex	L3	25
patent_intensity	Livestock AI patent intensity	OpenAlex	L3	25

Table A2. Pairwise Mann–Whitney U Tests with Bonferroni Correction (2022 AWPRI by Risk Tier)

Tier A	Tier B	U	p (uncorrected)	p (Bonferroni)	Significant
Critical	High	4	0.400	1.000	No
Critical	Moderate	11	0.167	1.000	No
Critical	Low	9	0.200	1.000	No
High	Moderate	44	0.001	0.009	Yes***
High	Low	36	0.003	0.017	Yes*
Moderate	Low	99	< 0.001	0.001	Yes***

Note. Critical tier n = 1 in k-means solution (China as singleton); comparisons involving Critical tier should be interpreted with caution. * $p < 0.05$; *** $p < 0.001$.