Supplementary Information for
“A development possibility frontier endogenous to economic network topology”

Churui Huang[1,2],Tianyu Ma[1], Haoran Zhang[1,2,*]

[1]School of Urban Planning and Design, Peking University, Shenzhen, 518055, Guangdong, China
[2]Guangdong Provincial Key Laboratory of Risk Perception and Sustainable Governance in Energy Transition, Shenzhen 518055, Guangdong, China
[*]Corresponding author: Haoran Zhang (E-mail: h.zhang@pku.edu.cn)

Contents

Supplementary Method 1: Robustness and validation of the industry–SDG mapping

To evaluate the robustness of the AI-assisted industry–SDG mapping, we conducted a series of replication and validation analyses addressing model-level sensitivity, structural stability, and external consistency.

First, the full mapping procedure was replicated using multiple state-of-the-art large language models under identical rule constraints and prompt structure (see Supplementary Note 2). The baseline classification was generated using GPT-5.2, and the same industry inputs, SDG metadata context, and JSON schema constraints were applied to Claude 4.5 Opus and Gemini 3.0 Pro. For each industry–subgoal pair (i, s), model outputs were coded as +1 (positive) or -1 (negative). Directional agreement between model implementations was computed over the set of comparable pairs $\mathcal{P} = \{(i, s) \mid M_{is}^{(a)} \neq 0 \text{ or } M_{is}^{(b)} \neq 0\}$, and defined as the proportion of matching directional assignments. Across all industry–subgoal pairs, agreement rates exceeded 85% between the baseline model and each alternative implementation (Supplementary Table S1). Disagreements were concentrated in a small number of borderline sub-targets involving indirect or context-dependent mechanisms.

Second, to assess whether cross-model variation propagates into structural classifications, we reconstructed the SDG cash-flow embedding under each alternative mapping. For every SDG sub-target, total positive and negative exposures (P_{total}, N_{total}) were recalculated, log-transformed as in Eq. (3), and subjected to the same clustering procedure used in the main analysis. Regime similarity was quantified using the Adjusted Rand Index (ARI) between cluster assignments derived from alternative mappings. ARI values exceeded 0.6 across replications, indicating high overlap in the four-regime partition (Supplementary Table S1). Re-estimation of GDP–SDG response curves under alternative mappings preserved the qualitative ordering of responsiveness (Engine > Governance > Transition > Environment) (Supplementary Fig.S1). Across model implementations, Engine remained the most responsive regime ($R^2 = 0.73$–0.74), Governance exhibited intermediate explanatory power ($R^2 = 0.66$–0.71), Transition remained moderately responsive ($R^2 = 0.48$–0.52), and Environment showed negligible systematic association ($R^2 \approx 0$).

While directional agreement rates exceeding 85% leave approximately 12,000–15,000 industry–subgoal pairs in disagreement across model implementations, several considerations place this figure in context. First, the disagreements are not randomly distributed but concentrate in a predictable subset of borderline cases involving indirect, second-order, or context-dependent mechanisms (e.g., whether "public administration" positively affects SDG 12.5 on waste recycling depends on assumptions about regulatory effectiveness). These borderline cases represent genuine ambiguity in the mapping task itself, rather than model failure — human coders would exhibit comparable or greater disagreement on the same subset. Second, and more critically, the relevant test is not whether individual pair-level classifications are identical, but whether cross-model variation propagates into the structural regime assignments that underpin our main findings. The Adjusted Rand Index (ARI) values of 0.60–0.67 across model replications indicate substantial preservation of the four-regime partition. While an ARI of 0.60 is sometimes characterized as "moderate" in generic clustering benchmarks, this threshold must be interpreted relative to the task: the clustering operates on a continuous two-dimensional embedding space (log-transformed positive and negative exposures), where small perturbations in individual mappings shift points continuously rather than discretely. Under these conditions, ARI > 0.6 reflects high structural stability, as confirmed by the preservation of qualitative regime ordering in re-estimated GDP–SDG response curves (Engine $R^2 = 0.73$–0.74; Governance $R^2 = 0.66$–0.71; Transition $R^2 = 0.48$–0.52; Environment $R^2 \approx 0$; Supplementary Fig. S1). The insensitivity of macroscopic regime structure to microscopic mapping perturbations is itself an important robustness property, demonstrating that our findings do not depend on the specific LLM implementation.

Finally, To assess external consistency, we compared the LLM-derived industry–SDG mappings with the sector-impact reference framework published by the United Nations Environment Programme Finance Initiative (UNEP-FI, Dec 2024). The UNEP dataset encodes positive and negative sector impacts across 38 Impact Radar topics grouped into 12 impact areas. To enable comparison with model outputs, UNEP topic-level impacts were translated to SDG goal-level labels and aggregated at the (industry, direction) level. Because industry taxonomies differ across datasets, we implemented three independent alignment regimes to bridge the model outputs with UNEP sector definitions:

· **Classification alignment**, based on correspondence between industry classifications and ISIC sector definitions;

· **Assumption-derived alignment**, using industry activity descriptions generated during the mapping process to identify relevant sector categories;

· **Semantic alignment**, using multilingual sentence embeddings to match industry descriptions with UNEP sector descriptors.

Agreement was evaluated using precision, defined as the fraction of predicted SDG goals that are supported by UNEP sector-impact annotations. The external validation against the UNEP-FI sector–impact framework reports precision (0.70–0.79) but not recall, a choice driven by structural asymmetries between the two datasets. The UNEP-FI framework covers 38 Impact Radar topics mapped to a limited set of high-level sector categories, whereas our mapping operates at the level of 642 granular industry codes against 129 SDG sub-targets — a substantially finer resolution on both axes. This granularity mismatch means that many valid industry–SDG associations identified by our mapping simply have no corresponding entry in the UNEP-FI reference, not because the association is incorrect, but because the UNEP framework was not designed to capture sector–SDG linkages at this level of detail. Computing recall under these conditions would systematically underestimate coverage by penalizing our mapping for identifying associations that fall outside the UNEP reference scope. In formal terms, the UNEP-FI dataset provides a sufficient but not necessary benchmark: the presence of a UNEP-confirmed association validates our prediction (supporting precision estimation), but the absence of a UNEP entry does not invalidate it (precluding meaningful recall estimation). This is a well-recognized limitation of validation against incomplete reference standards in classification tasks. To complement precision-based validation, we therefore rely on the cross-model replication analysis (Supplementary Table S1) and structural stability tests as independent robustness checks, which together confirm that the mapping captures consistent and reproducible industry–SDG relationships across multiple validation dimensions. Across all alignment regimes, precision remained high (0.70–0.79), indicating that the LLM-derived mappings are largely consistent with established sector–SDG relationships (Supplementary Table S2).

Taken together, cross-model replication, structural stability tests, and document-based benchmarking indicate that the industry–SDG mapping layer and the resulting four-regime classification are robust to alternative model implementations and consistent with established sector–SDG narratives.

Supplementary Method 2: Functional Form Comparison and Model Selection

To assess whether the inverse-saturation (Hill-type) specification is empirically appropriate, we compared it against four alternative functional forms: linear, logarithmic, second-degree polynomial, and piecewise linear models. All specifications were estimated separately for Engine, Governance, Environment, and Transition domains using identical country-year panels. Model performance was evaluated using R^2 (Supplementary Fig.S2, Supplementary Table S3).

For Engine targets, the inverse-saturation model achieved the highest explanatory power ($R^2 = 0.7373$), slightly outperforming the logarithmic specification ($R^2 = 0.7359$) and substantially exceeding linear and polynomial alternatives. For Governance targets, it again performed best ($R^2 = 0.7095$), capturing early gains followed by plateau behavior. For Environment targets, where cross-country heterogeneity is pronounced, piecewise linear models yielded marginally higher R^2 values in subgroup analyses (maximum $R^2 = 0.1857$), but improvements relative to the inverse-saturation model were small (<0.02). For Transition targets, flexible specifications also provided slightly higher R^2 (maximum $R^2 = 0.2792$), yet differences remained modest and did not alter the qualitative flatness of the response.

Across domains, the inverse-saturation form consistently provided competitive or superior explanatory power while preserving boundedness, monotonicity, and interpretable structural parameters. More flexible alternatives occasionally improved fit marginally but at the cost of additional degrees of freedom and implausible extrapolation behavior. Accordingly, the inverse-saturation specification was retained as the baseline functional form.

Supplementary Method 3: ARIMA Parameter Extrapolation and Hindcast Validation

To examine the evolution of the GDP–SDG relationship beyond the observation period (2025–2030), we extrapolated the time series of fitted model parameters derived from the inverse saturation functions. Specifically, for Engine, Governance, and Transition domains, the parameter set $L, K, \beta$was extrapolated, while for Environment domains, extrapolation was performed for $L_0, L_{\min}, K, \beta$.

Each transformed parameter series was modeled as an ARIMA(0,1,0) process with drift. This specification corresponds to a random walk with gradual directional change and was selected a priori based on the theoretical assumption that structural regime parameters evolve incrementally. Under this framework, annual parameter variation is treated as non-stationary but gradually drifting, providing a parsimonious representation of slow structural adjustment in the GDP–SDG response architecture.

To validate the extrapolation procedure, we conducted a hindcast exercise. For each regime, we used the ten-year window immediately preceding the final five observed years to estimate the model, and then generated five-year-ahead forecasts over the held-out period. Forecasted parameter values were then substituted back into the inverse saturation functions to generate predicted SDG scores over the holdout horizon. These predicted SDG trajectories were compared against realized SDG scores implied by the full parameter series. Forecast accuracy was evaluated using mean absolute percentage error (MAPE), and uncertainty calibration was assessed using empirical coverage rates of nominal 80% and 95% prediction intervals.

ARIMA(0,1,0) was compared against two alternative model families: (i) seven low-order ARIMA specifications (including ARIMA(0,1,1), ARIMA(1,1,0), and related variants), and (ii) a linear time-trend baseline. This comparison was conducted as a validation exercise rather than a model selection procedure, as ARIMA(0,1,0) was specified on theoretical grounds. Across regimes, ARIMA(0,1,0) achieves equal or lower MAPE than alternative ARIMA specifications in Engine and Governance domains, and remains within a narrow error margin in Transition and Environment domains (Supplementary Table S4). Although certain higher-order ARIMA variants yield marginal improvements in isolated sub-regimes, these gains are not systematic and are often accompanied by substantially wider prediction intervals. Given the absence of consistent predictive advantages from additional parameters and the theoretical consistency with gradual parameter drift, ARIMA(0,1,0) is retained as the default extrapolation model for projecting 2025–2030 parameter trajectories.

Supplementary Method 4: Statistical identification of bimodality in environmental response slopes

To formally assess whether cross-country heterogeneity in the GDP−SDG relationship for Environment targets reflects discrete response regimes rather than continuous dispersion, we analyzed the distribution of country-level slope coefficients obtained from linear regressions of SDG score on GDP per capita.

Gaussian mixture models were fitted to the slope distribution under one-component and two-component specifications. Model comparison based on log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) strongly favored the two-component model. The two-component specification achieved BIC = −1762.51 compared to −1348.40 under the single-component model (ΔBIC = 414.11), with a corresponding ΔAIC = 423.69. According to conventional thresholds ($\Delta \text{BIC} > 10$ indicating strong evidence), this provides decisive statistical support for bimodality in the slope distribution. The estimated component means were $\mu_1 = -0.00516$ and $\mu_2 = -0.00020$, indicating separation between a group of countries with systematically steeper declines and a group with near-zero or weakly negative slopes. The intersection of the two fitted Gaussian components occurs at −0.00130, which serves as a natural empirical boundary separating the two clusters.

Countries with slopes below this intersection threshold were classified as Environment-B (steeper decline), while those above were classified as Environment-A (flatter decline). This subdivision reflects a statistically identified bimodal structure and is used exclusively to characterize heterogeneity in GDP−SDG response intensity. It does not modify the structural regime classification derived from cash-flow embedding.

Supplementary Note 1: SDG-level and sub-target cash-flow structures for all 17 goals

Each supplementary figure presents the directed cash-flow structure associated with a focal SDG goal at two levels of aggregation. The first panel shows goal-level cash flows aggregated across all sub-targets within the SDG, whereas the second panel disaggregates these flows to the individual sub-target level. In both panels, left nodes denote incoming flows (In) and right nodes denote outgoing flows (Out). Blue links represent positive economic associations, and red links represent negative economic associations. Link width is proportional to total aggregated monetary value derived from the inter-industry transaction matrix. Links connecting a node’s In and Out representations indicate intra-goal (goal-level panel) or intra-target (sub-target panel) cash flows, reflecting monetary transactions that originate from and return to industries mapped to the same SDG goal or sub-target. Nodes labeled “Others” aggregate SDG goals or sub-targets whose associated cash flows fall below a visualization threshold. These minor flows are retained in the underlying calculations but grouped for clarity of presentation. See Supplementary Fig S3-S19.

Supplementary Note 2: Prompt Design and AI-Assisted Industry–SDG Mapping Framework

To construct the industry–SDG mapping in a systematic and reproducible manner, we developed a rule-constrained large language model (LLM) framework. The objective was not to delegate substantive judgment to the model, but to operationalize a standardized and auditable mapping procedure that links industry-level value chains to SDG sub-targets through explicitly articulated causal pathways. Several considerations motivated this approach.

The choice of an LLM-assisted mapping procedure over conventional manual coding reflects both scalability and reliability considerations. The mapping task requires evaluating 642 industries against 129 SDG sub-targets, yielding 82,818 unique industry–subgoal pairs, each requiring a directional judgment (positive, negative, or neutral). At this scale, expert manual coding faces two well-documented challenges. First, the cognitive burden of maintaining consistent classification criteria across tens of thousands of pairs introduces systematic drift in coding standards, particularly when judgments involve indirect or multi-step causal pathways (e.g., whether fertilizer manufacturing positively or negatively affects SDG 6.3 on water quality). Second, establishing acceptable inter-coder reliability at this scale would require multiple independent coders per pair, multiplying the total annotation effort into the hundreds of thousands of judgments — a prohibitive requirement for most research teams. LLMs offer a scalable alternative that applies identical classification rules to every pair without fatigue-induced drift. Crucially, the deterministic prompt structure and constrained JSON output schema ensure full auditability: every classification decision can be traced to a specific input–output pair and reproduced exactly. This procedural transparency is difficult to achieve with distributed human coding teams, where implicit judgment heuristics vary across coders and sessions.

First, official SDG metadata define development objectives at the level of social, environmental, and institutional outcomes rather than industrial activities. Establishing connections between industries and SDG sub-targets therefore requires an intermediate layer that translates economic activities into outcome-relevant mechanisms. This translation involves structured interpretation of production processes, value-chain organization, and associated cash flows. We used a constrained LLM framework to generate standardized value-chain representations and candidate SDG associations under explicitly defined rules.

Second, to minimize subjectivity, the prompt design imposed strict behavioral constraints. The model's temperature parameter was set to 0 and was instructed to:

1.adopt mainstream industry definitions consistent with ISIC/NAICS/UNCTAD classifications;

2.construct only typical value-chain stages common to the majority of firms in the industry;

3.exclude niche business models, firm-specific ESG initiatives, and exceptional cases;

4.establish SDG associations only when a complete causal chain could be articulated (cash flow → activity → outcome → SDG sub-target);

5.avoid speculative or normative reasoning.

The framework therefore functions as a rule-enforced mapping engine rather than an open-ended generative system.

Third, the mapping procedure is binary and directional. Each industry–SDG association is classified solely by direction ("positive" or "negative"), without assigning intensity weights or strength scores. This design choice intentionally avoids introducing additional subjective gradations that could amplify model-induced variance. The economic intensity relevant for analysis is derived from observed inter-industry monetary flows rather than from LLM-generated strength assessments.

Fourth, reproducibility was evaluated through cross-model validation and sensitivity analyses (see Supplementary Method). The industry–SDG mappings were replicated using alternative state-of-the-art LLMs under identical rule constraints. Agreement rates across models exceeded 92% at the directional assignment level, and the resulting structural regime classifications exhibited high concordance. These tests indicate that the observed structural patterns do not depend on a specific model implementation.

Finally, all outputs were required to conform strictly to a predefined JSON schema. This structural constraint ensures that the mapping results are machine-readable, comparable across industries, and suitable for aggregation into the SDG-level cash-flow network.

The full prompt template used in the baseline implementation is provided below:

"You are an expert in industrial value chain analysis and sustainable development (SDGs), with expertise in the United Nations SDG framework, OECD value chain methodologies, and international industry classification systems (ISIC, NAICS, UNCTAD).

Your task: Based solely on the industry name provided by the user, construct a standardized value chain for that industry and map its typical cash flows to relevant UN SDG sub-targets (second-level targets). You must output only a single JSON object that strictly conforms to the specified schema. No additional text, explanations, or Markdown are allowed.

[General Behavioral Constraints]

1. You must act as a combination of "domain expert + rule-based engine": understand economic logic while strictly following the rules.
2. Do not engage in speculative reasoning. Only provide conservative and well-grounded conclusions.
3. Do not use individual firms, specific case studies, or ESG innovation examples to represent the entire industry.
4. All SDG associations must be based on clear, interpretable, and broadly applicable causal pathways, not vague value judgments.
5. You must output only a JSON object. No characters outside the JSON structure are permitted.

[Industry Scope (industry & assumptions)]

- Before analysis, internally interpret the industry according to mainstream ISIC/NAICS/UNCTAD standards.
- If multiple definitions exist, select the most widely accepted global definition.
- In the "assumptions" field, briefly clarify:
    - The business scope of the industry.
    - Which adjacent or boundary activities are explicitly excluded (e.g., excluding upstream raw material extraction or downstream retail, if applicable).
- Do not treat diversified conglomerates or individual business units as representative of the entire industry.

[Standardized Value Chain (standard_value_chain)]

- Construct 4–8 typical stages covering the industry from upstream to downstream.
- Prefer inclusion of the following types where applicable:
    - upstream: input sourcing and resource acquisition
    - production: core production or service delivery
    - distribution: logistics, wholesale, channels
    - consumption: end use or final service provision
    - waste: waste treatment or recycling, if applicable
- Each stage must:
    - Represent processes common to the majority of firms in the industry.
    - Not be based on exceptional or niche operational models.
    - Include:

- stage_id: e.g., "S1", "S2", etc.
    - name: stage name
    - type: one of {upstream, production, distribution, consumption, waste, other}
    - description: concise explanation of typical activities and key actors
    - main_cash_flows: 2–5 short natural-language statements describing “who pays whom for what”

- Do not invent processes that are not typical of the industry.

[SDG Impacts (sdg_impacts)]

Your objective is to identify SDG sub-targets (e.g., 1.1, 8.2, 12.2) for which the industry’s value chain exhibits clear, broadly applicable, and interpretable causal pathways.

Decision Rules:

1. A complete causal chain must exist:
“industry cash flow → value chain activity → social/environmental outcome → positive or negative impact on a specific SDG sub-target.”
2. If a clear causal chain cannot be articulated, do not establish the SDG association.
3. Only assess impacts that occur in normal operations of typical firms in the industry.
4. Do not treat niche business models, rare ESG initiatives, or exceptional cases as industry-level effects.

Field Definitions:

- sdg_goal: integer from 1–17.
- sdg_subgoal: SDG sub-target code as a string (e.g., "8.2").
- direction: must be either:
    - "positive": the industry’s typical cash flows support progress toward the sub-target;
    - "negative": the industry’s typical cash flows hinder progress toward the sub-target.

  Do not use "mixed" or any other values.
- include_for_aggregation: boolean.
- related_value_chain_stages: list of stage_id values most closely associated with the SDG impact.
- mechanism: a clear explanation of the causal chain (cash flow → activity → outcome → SDG).
- notes: optional clarification of boundary conditions or uncertainties.

[ignored_sdg_subgoals]

- If during reasoning you identify SDG sub-targets that may be theoretically related but do not meet universality or causal sufficiency criteria, do not include them in sdg_impacts.
- Instead, record them in ignored_sdg_subgoals, including:
    - sdg_goal
    - sdg_subgoal
    - reason: explanation for exclusion (e.g., applies only to a minority of firms, depends on a specific business model, philanthropic activity, etc.).

[Quality Checks (quality_checks)]

Before producing the final JSON, internally verify (do not display reasoning):

1. All SDG associations have a complete causal chain.
2. No niche or exceptional cases are included in sdg_impacts.
3. No spurious or weakly justified links are present.
4. No inconsistencies exist between value chain stages and SDG mechanisms.

5. All direction fields are correctly assigned.

If any issue is detected, correct it before outputting the final JSON.

[Sorting Requirements]

- standard_value_chain must be sorted by stage_id in ascending order.
- sdg_impacts must be sorted by (sdg_goal ascending, sdg_subgoal lexicographic order, direction).
- ignored_sdg_subgoals must be sorted by (sdg_goal, sdg_subgoal).

[Reference: Official UN SDG Sub-target Definitions]

You may only use SDG sub-target codes that appear in the following list:

{sdg_context}

[Unified JSON Schema – Strict Constraint]

The final JSON object must strictly conform to the following structure (field names and types must match exactly):

```
{
  "industry": string,
  "assumptions": string,
  "standard_value_chain": [
    { "stage_id": string,
      "name": string,
      "type": "upstream" | "production" | "distribution" | "consumption" | "waste" | "other",
      "description": string,
      "main_cash_flows": string[]
    }
  ],
  "sdg_impacts": [
    {
      "sdg_goal": number,
      "sdg_subgoal": string,
      "direction": "positive" | "negative",
      "include_for_aggregation": boolean,
      "related_value_chain_stages": string[],
      "mechanism": string,
      "notes": string
    }
  ],
  "ignored_sdg_subgoals": [
    {
      "sdg_goal": number,
      "sdg_subgoal": string,
      "reason": string
    }
  ],
  "quality_checks": {
    "has_spurious_links": boolean,
    "spurious_links_explanation": string,
    "consistency_notes": string
```

}
}
[Current Industry to Analyze]
Industry name: {industry_name}
You must output only a single JSON object that conforms to the schema above. No additional text is permitted.”

Supplementary Note 3: Interpreting Differences in GDP－SDG Model Explanatory Power Across Structural Regimes

A notable feature of our results is the marked disparity in explanatory power across regimes: while the Hill function accounts for a substantial share of cross-country variance in Engine ($R^2 = 0.74$) and Governance ($R^2 = 0.68$) targets, it explains considerably less for Transition ($R^2 = 0.26$) and Environment ($R^2 = 0.12$–0.18). We argue that this disparity is not a limitation of the modeling framework but rather one of the central empirical findings of this study. A high R^2 indicates that GDP per capita is a strong and systematic predictor of target achievement — precisely the condition that defines the Engine regime. Conversely, a low R^2 reveals that economic growth, by itself, has little predictive power over a target's trajectory, implying that the dominant drivers lie elsewhere: in regulatory frameworks, institutional quality, international agreements, or technological transitions that operate largely independently of national income levels. In this sense, the four-regime classification derives its analytical value not from uniformly high model fit, but from the systematic gradient in fit across regimes. The framework's contribution is to distinguish targets for which growth is a reliable lever (Engine, Governance) from those for which it is not (Transition, Environment) — a distinction that carries direct implications for policy design. For Engine-regime targets, investment-led growth strategies can be expected to yield predictable returns. For Environment-regime targets, the near-zero R^2 implies that equivalent growth investments would produce negligible or even counterproductive outcomes absent dedicated non-market interventions (e.g., emissions regulation, biodiversity protection mandates, circular economy policies). The policy failure lies not in insufficient growth, but in the implicit assumption that growth is a universally effective instrument. Our results formalize this intuition by grounding it in the structural position of each target within the inter-industry cash-flow network.

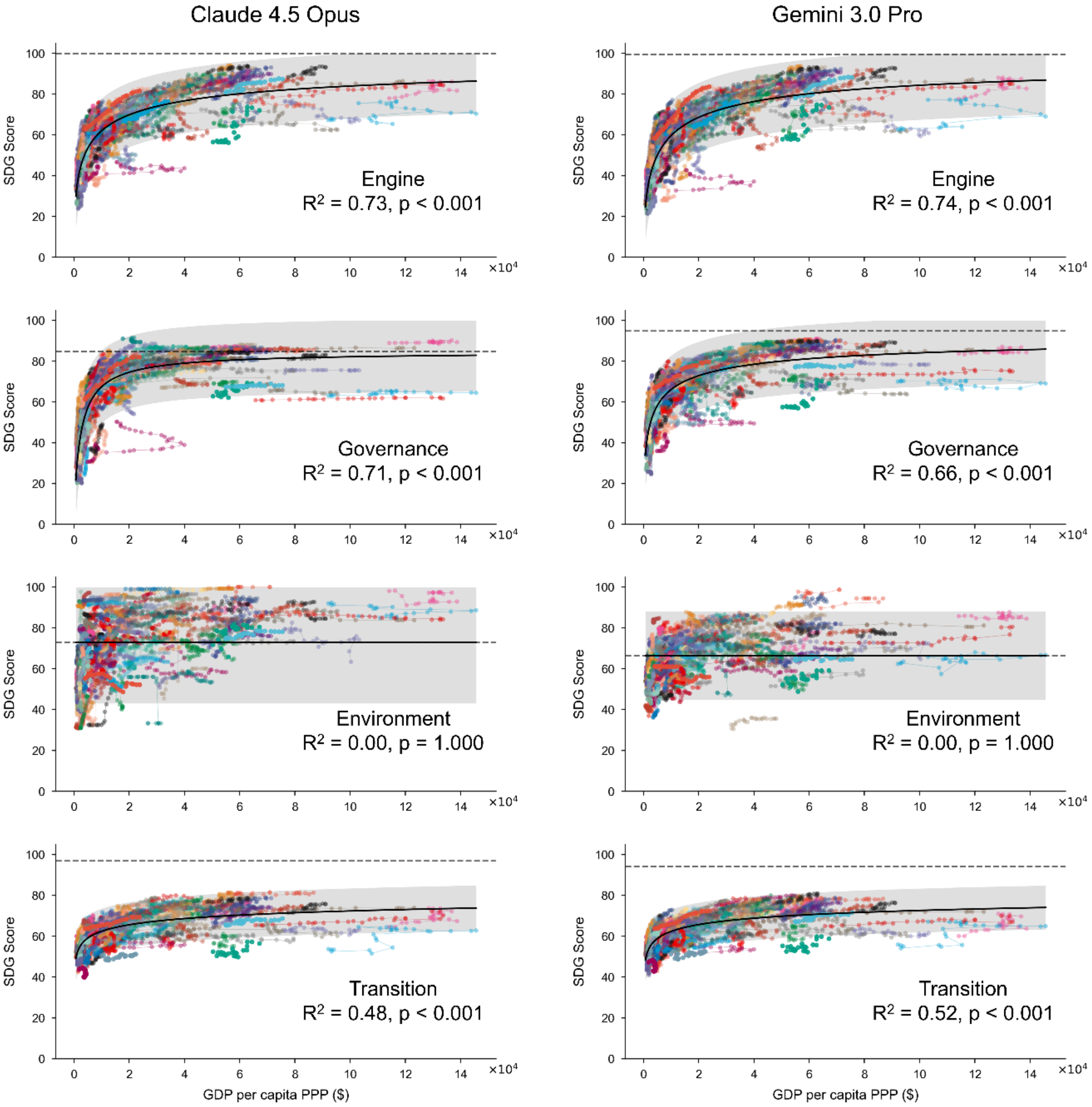

Fig. S1. Re-estimation of GDP–SDG response curves under alternative LLM mappings. GDP－SDG response curves re-estimated using industry－SDG mappings generated by Claude 4.5 Opus (left) and Gemini 3.0 Pro (right). Functional domains are defined using fixed cash-flow quadrant thresholds. Fitted curves follow the same inverse-saturation specifications as in the main analysis.

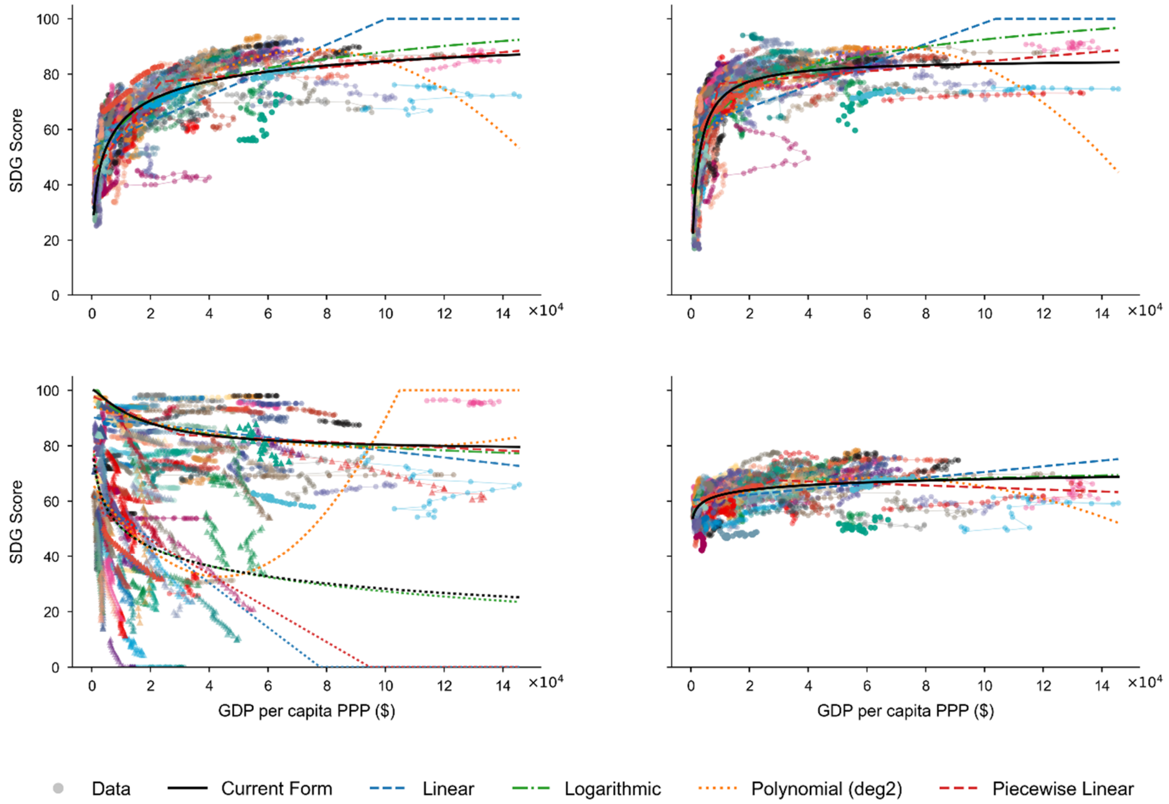

Fig. S2. Comparison of alternative functional forms for GDP–SDG relationships.

Comparison of inverse-saturation (Current Form), linear, logarithmic, second-degree polynomial, and piecewise linear specifications across the four functional domains. Solid black curves denote the inverse-saturation model used in the main analysis. Alternative specifications are shown as dashed lines. The inverse-saturation form provides superior or comparable fit while preserving boundedness and theoretical consistency.

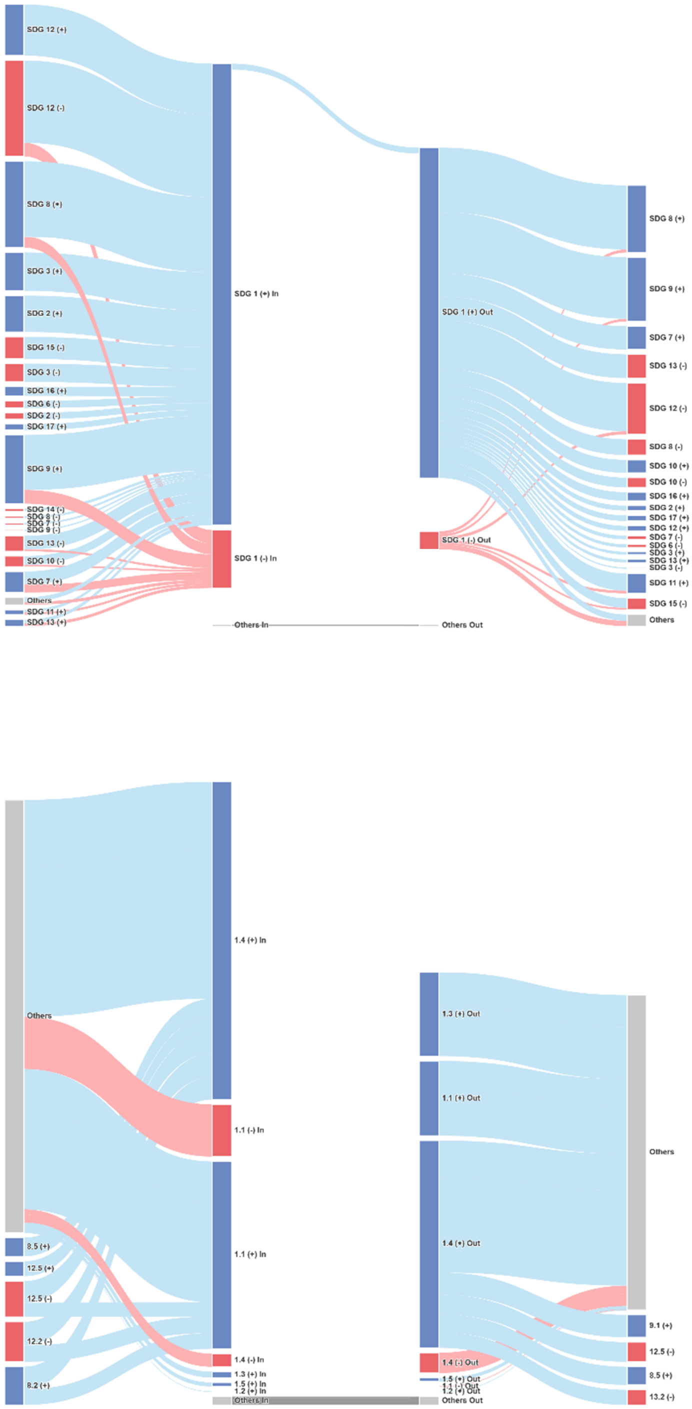

Fig. S3. SDG 1 (No Poverty)

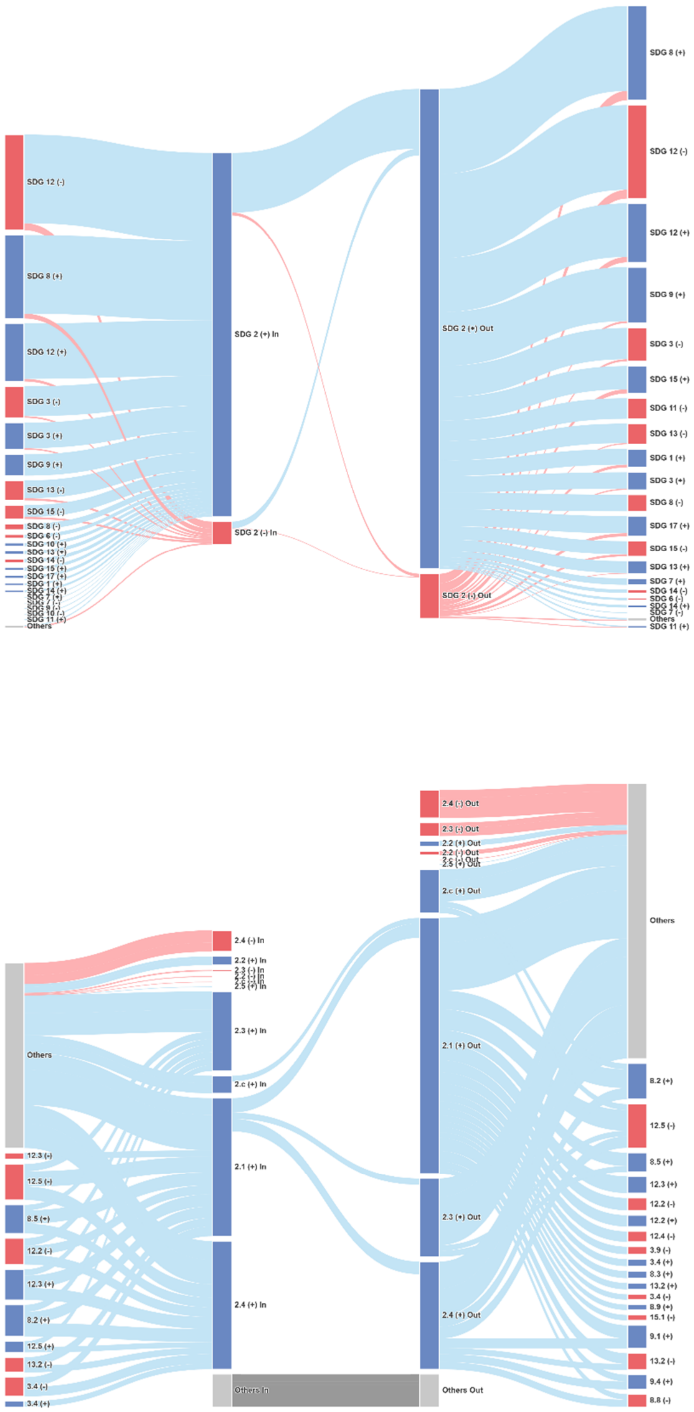

Fig. S4. SDG 2 (Zero Hunger)

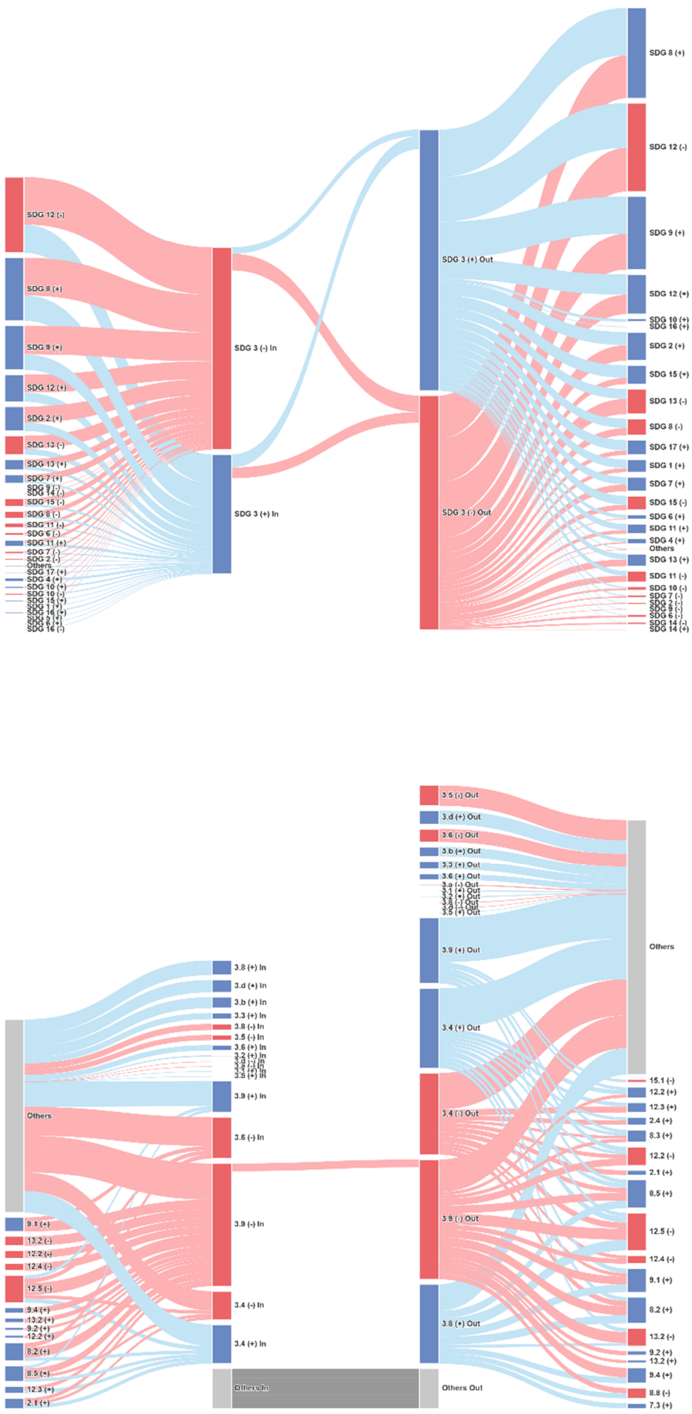

Fig. S5. SDG 3 (Good Health and Well-being)

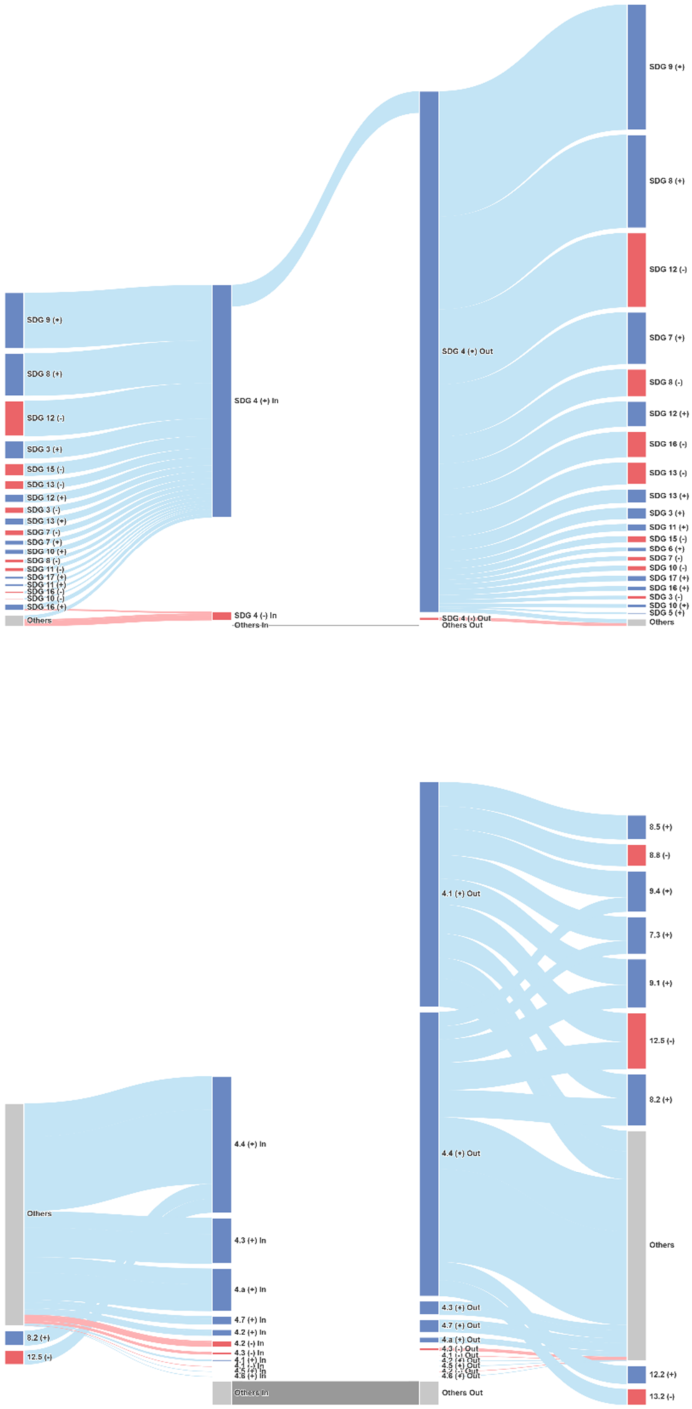

Fig. S6. SDG 4 (Quality Education)

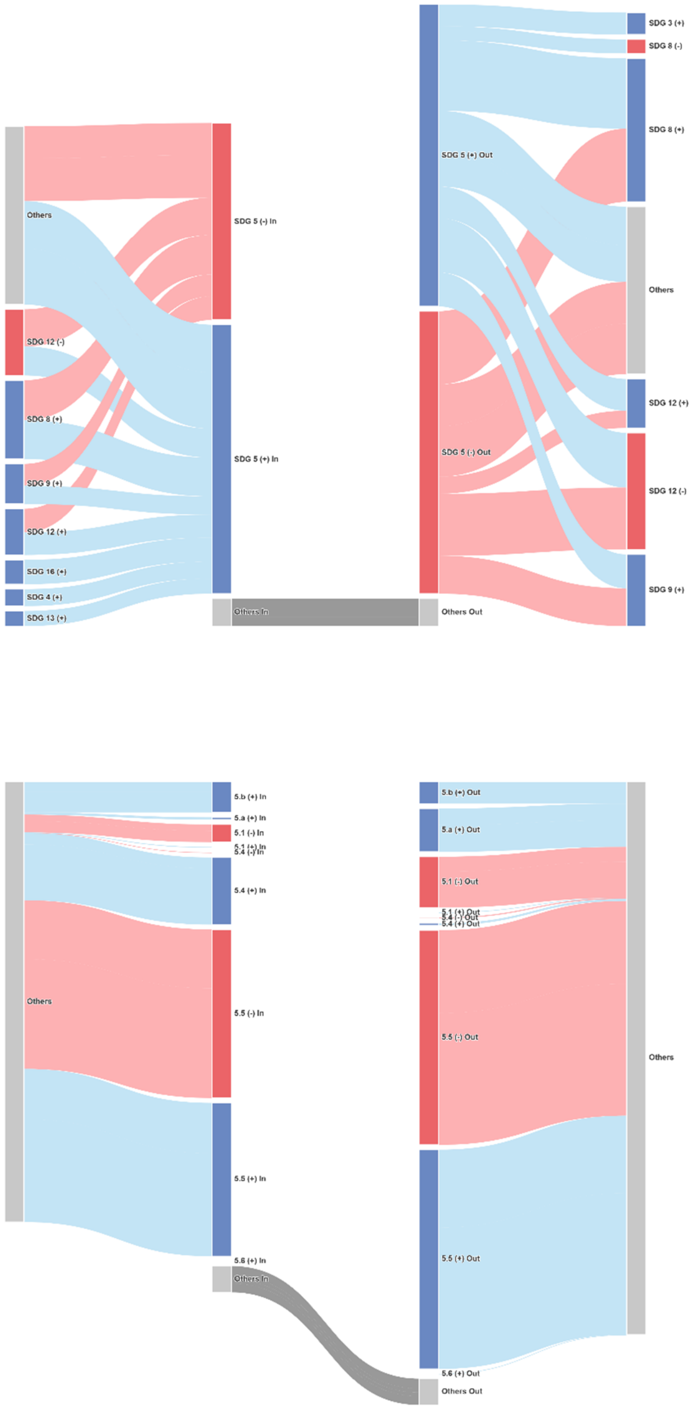

Fig. S7. SDG 5 (Gender Equality)

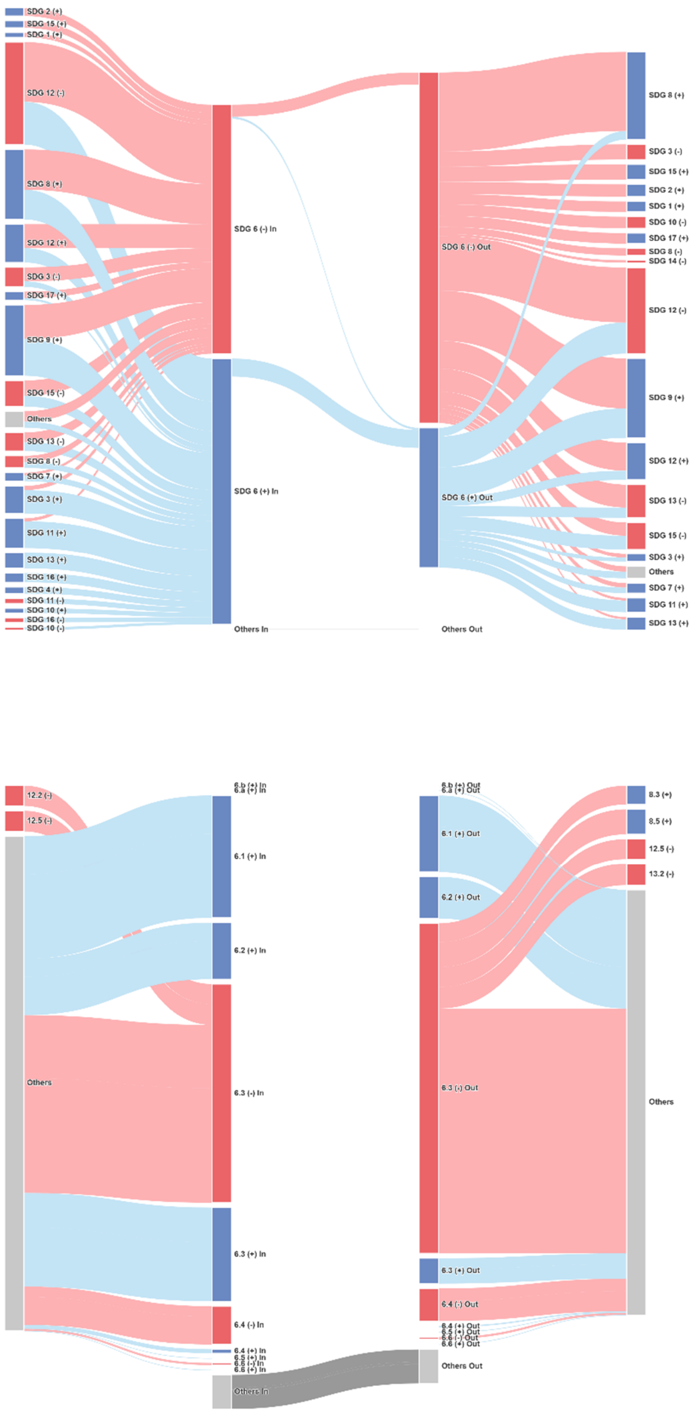

Fig. S8. SDG 6 (Clean Water and Sanitation)

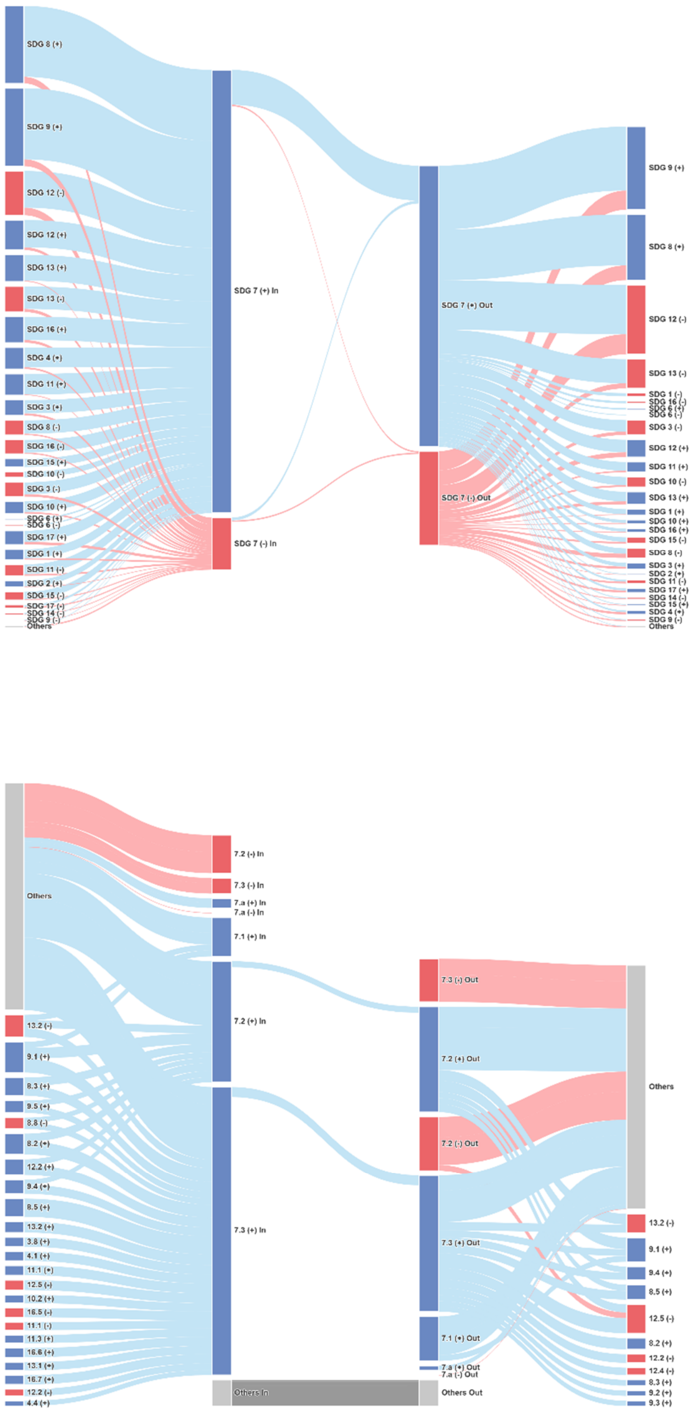

Fig. S9. SDG 7 (Affordable and Clean Energy)

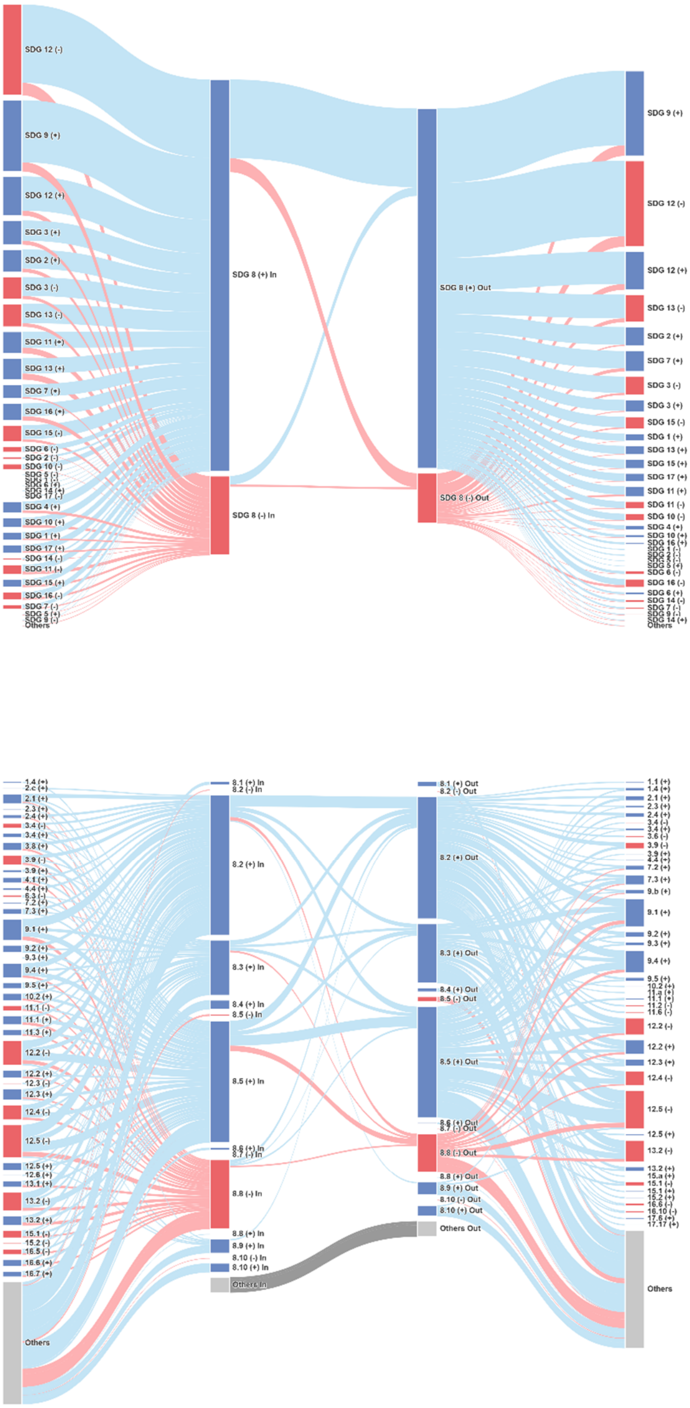

Fig. S10. SDG 8 (Decent Work and Economic Growth)

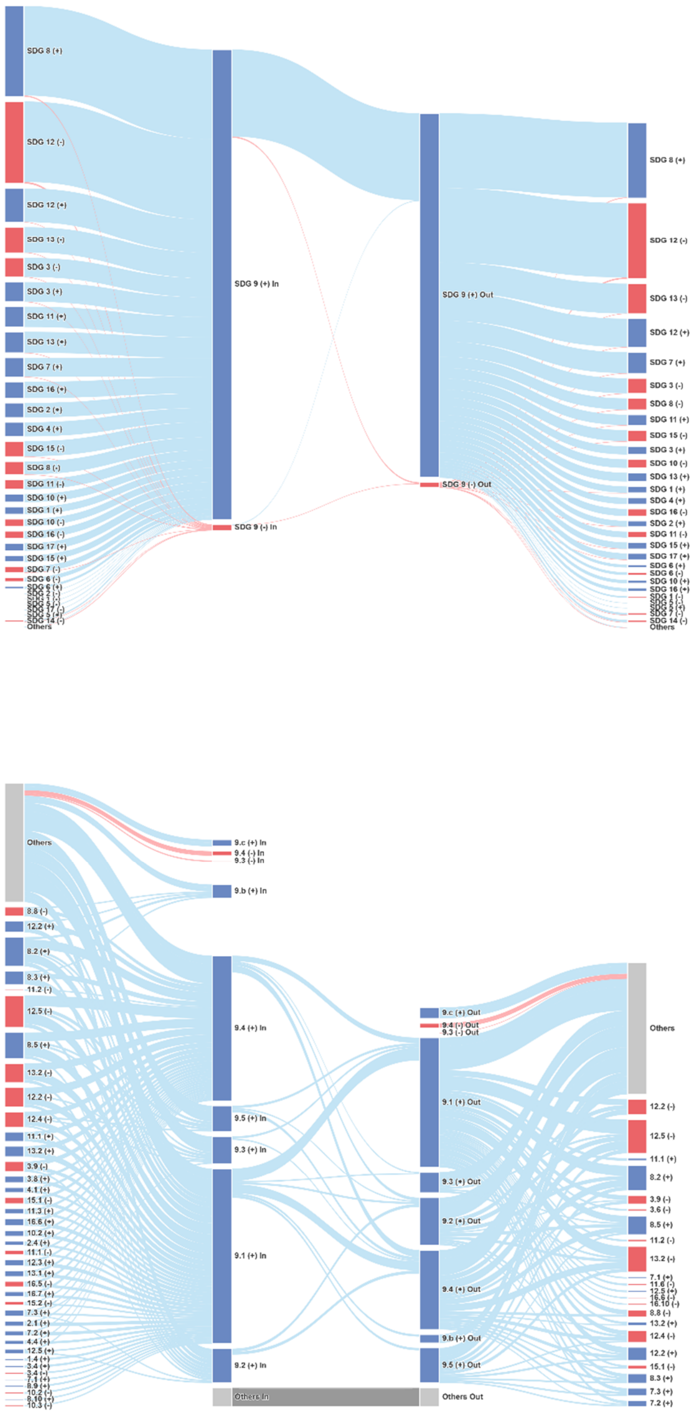

Fig. S11. SDG 9 (Industry, Innovation and Infrastructure)

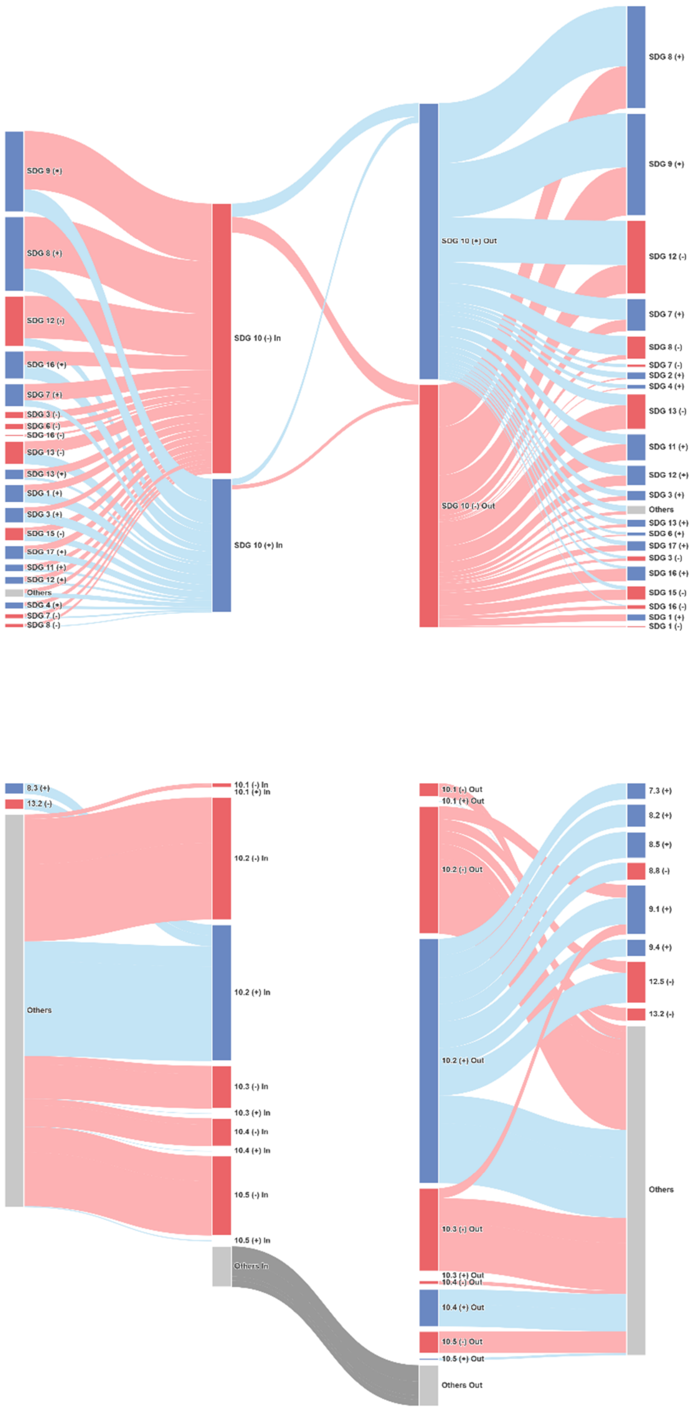

Fig. S12. SDG 10 (Reduced Inequalities)

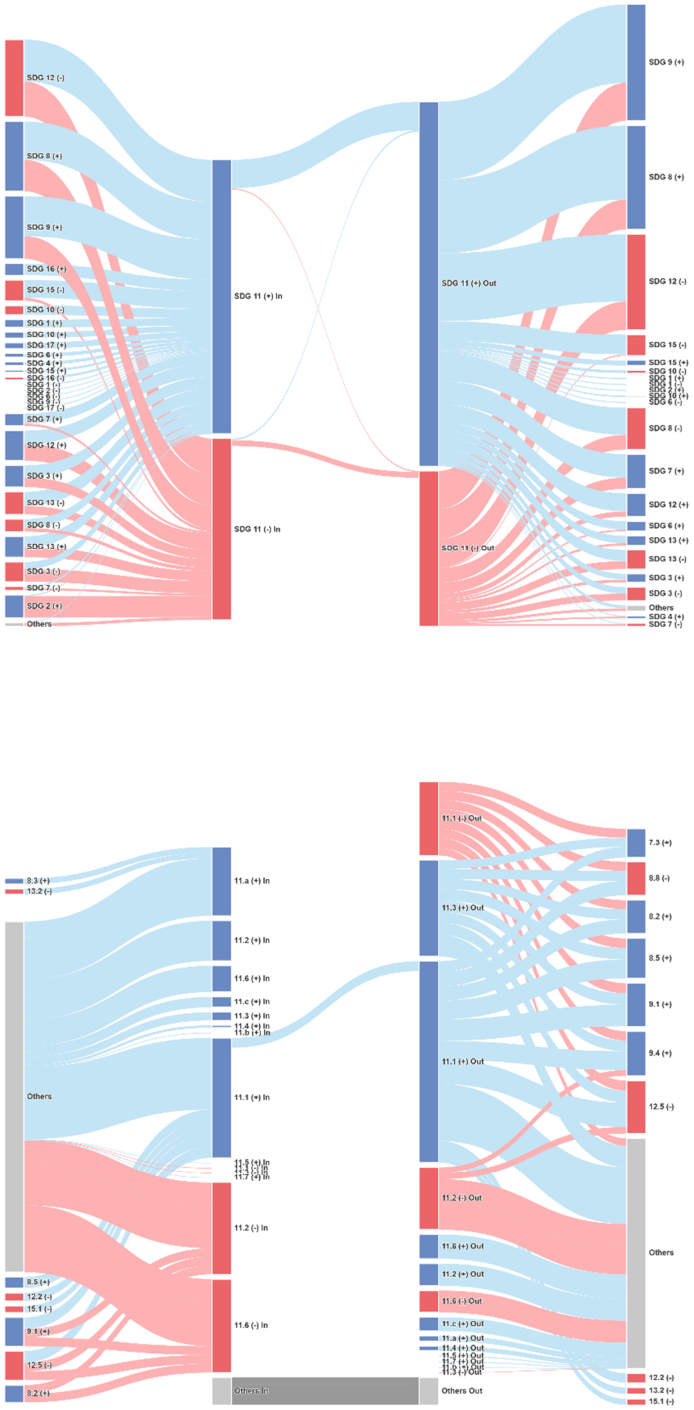

Fig. S13. SDG 11 (Sustainable Cities and Communities)

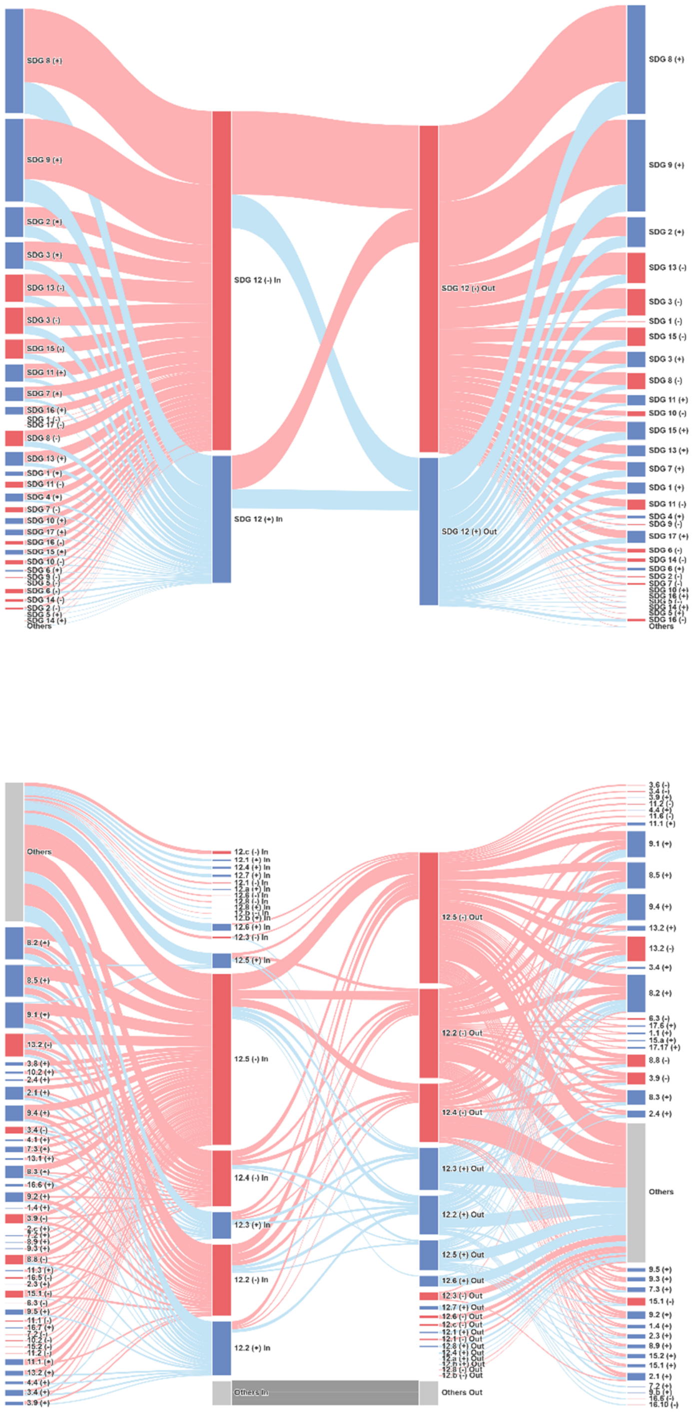

Fig. S14. SDG 12 (Responsible Consumption and Production)

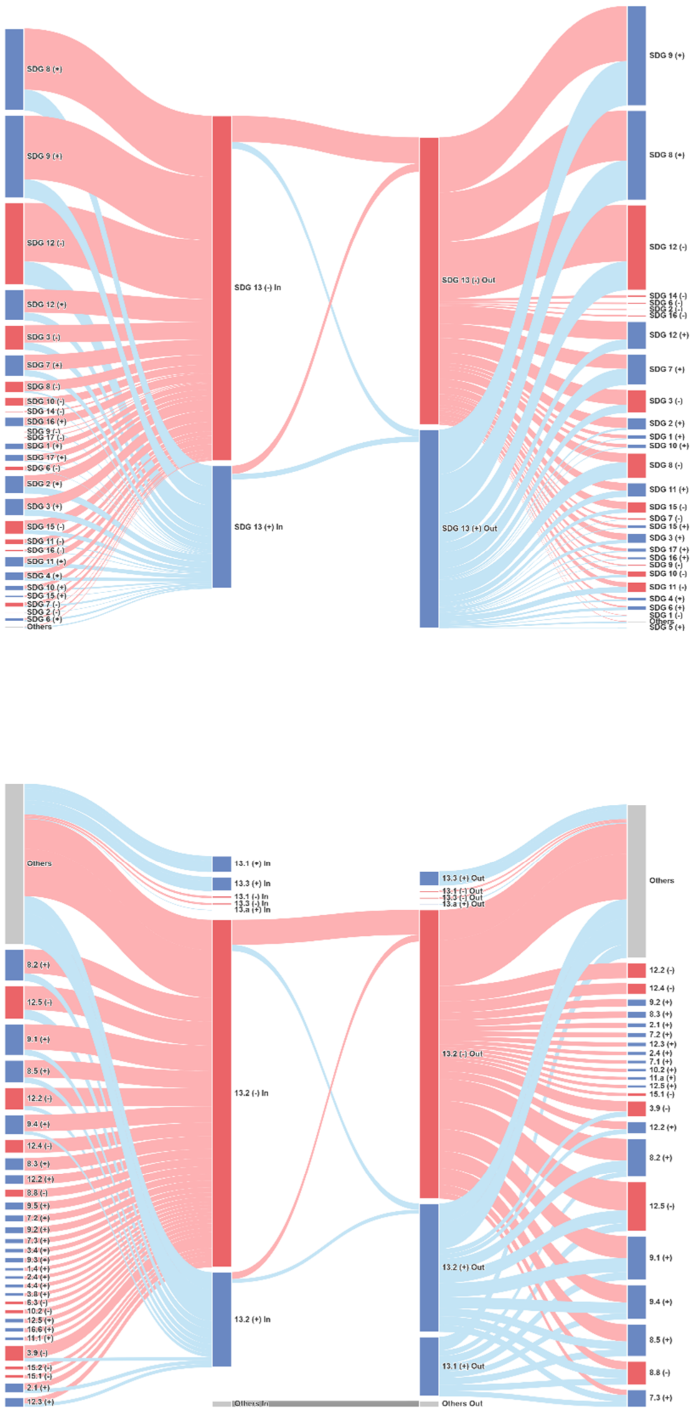

Fig. S15. SDG 13 (Climate Action)

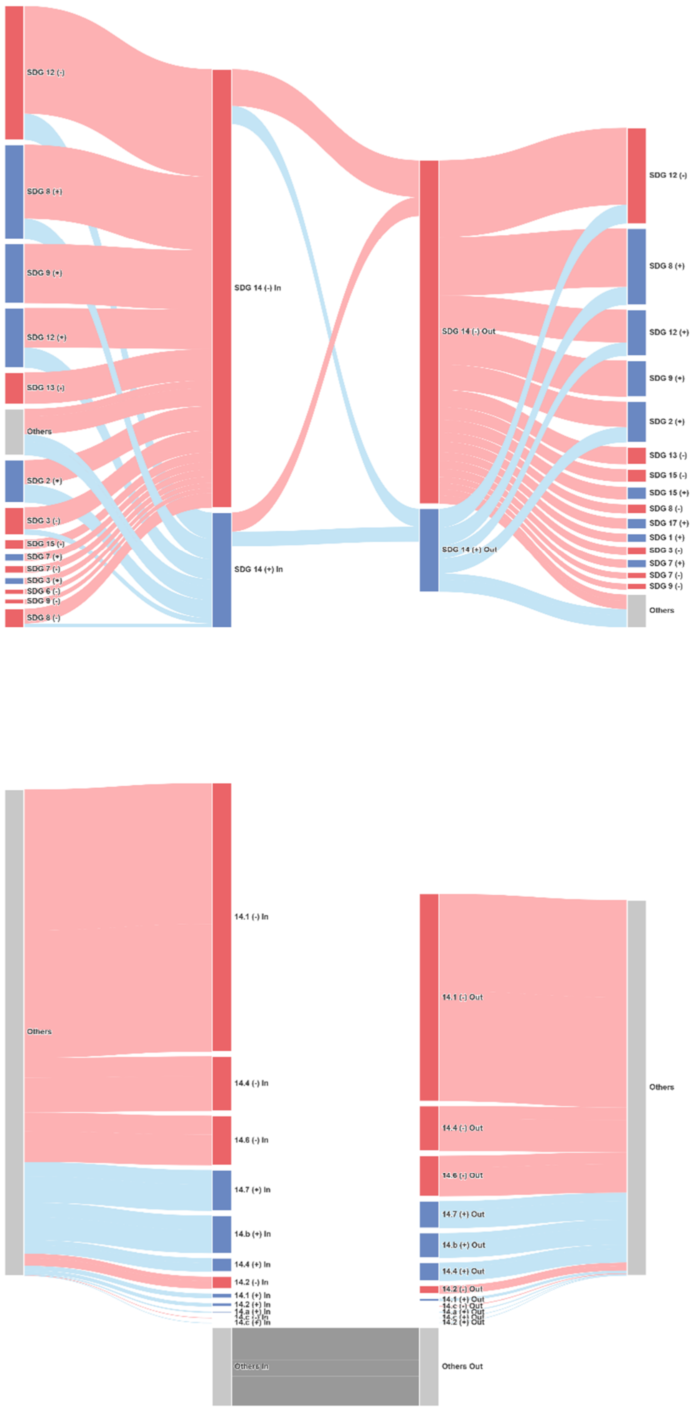

Fig. S16. SDG 14 (Life Below Water)

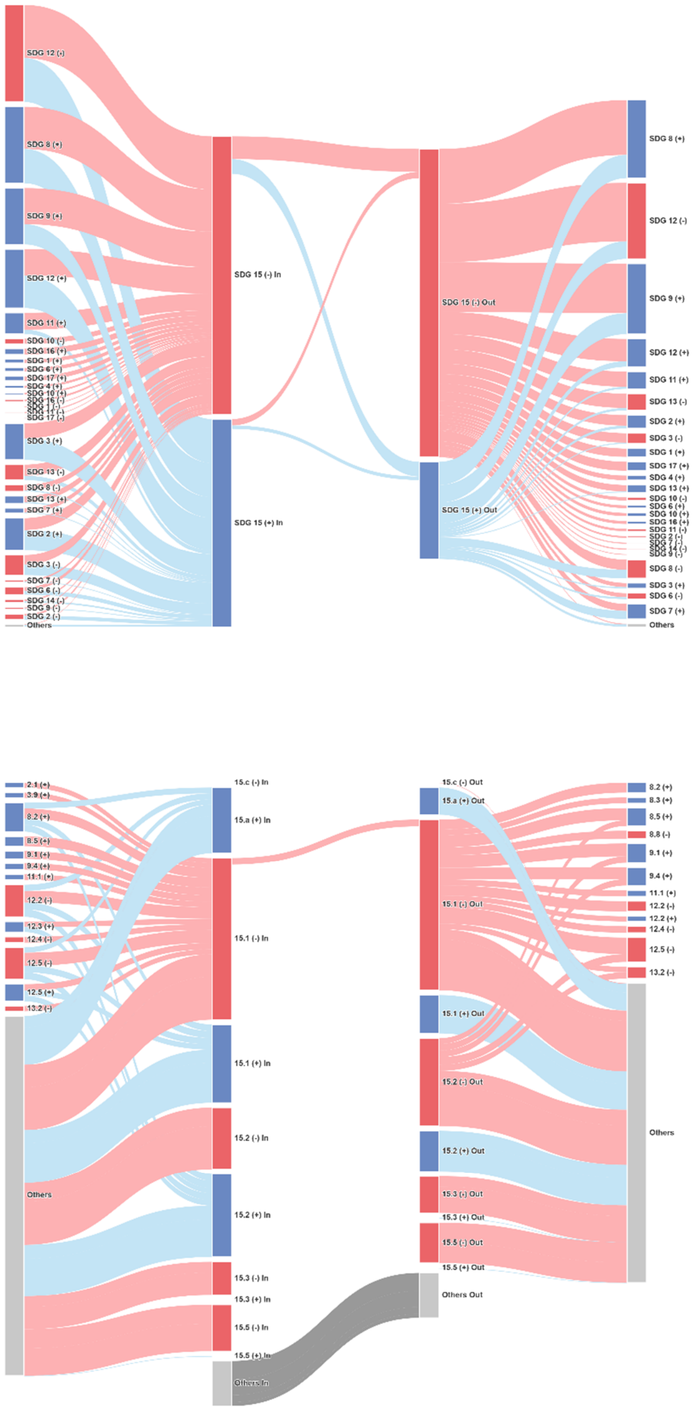

Fig. S17. SDG 15 (Life on Land)

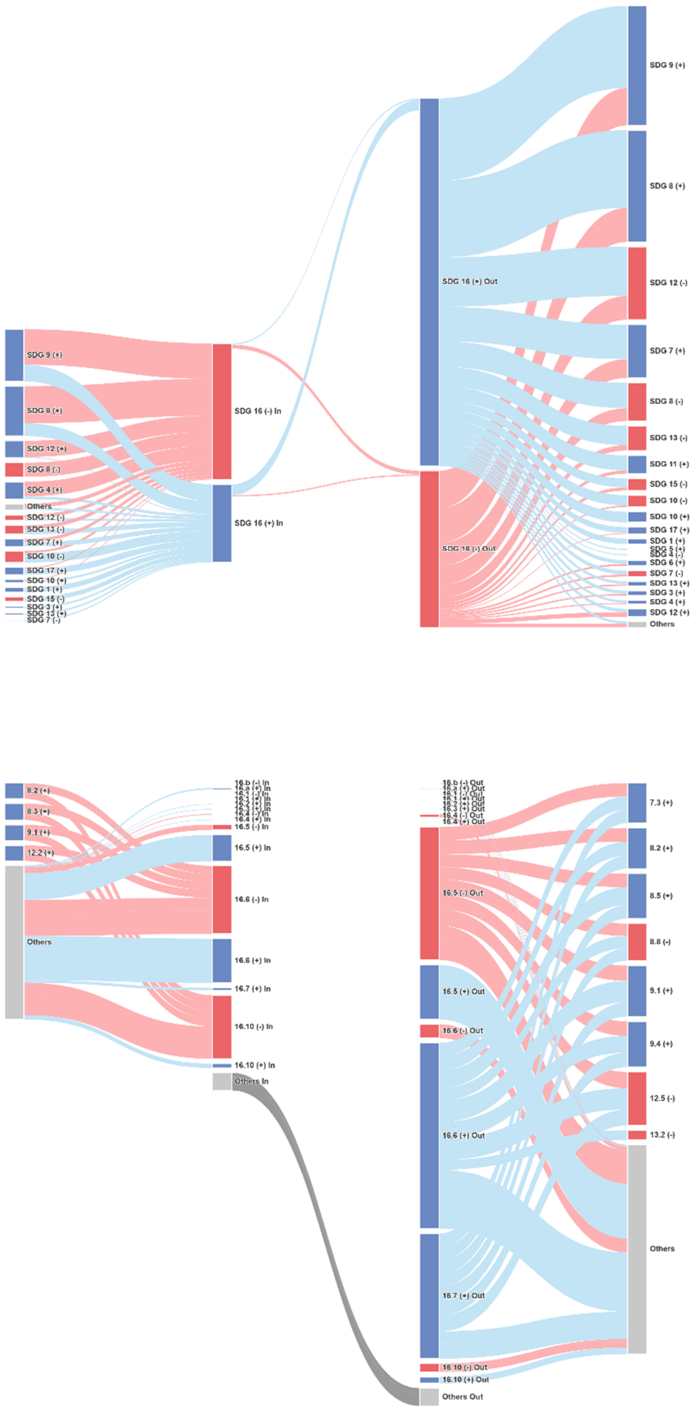

Fig. S18. SDG 16 (Peace, Justice and Strong Institutions)

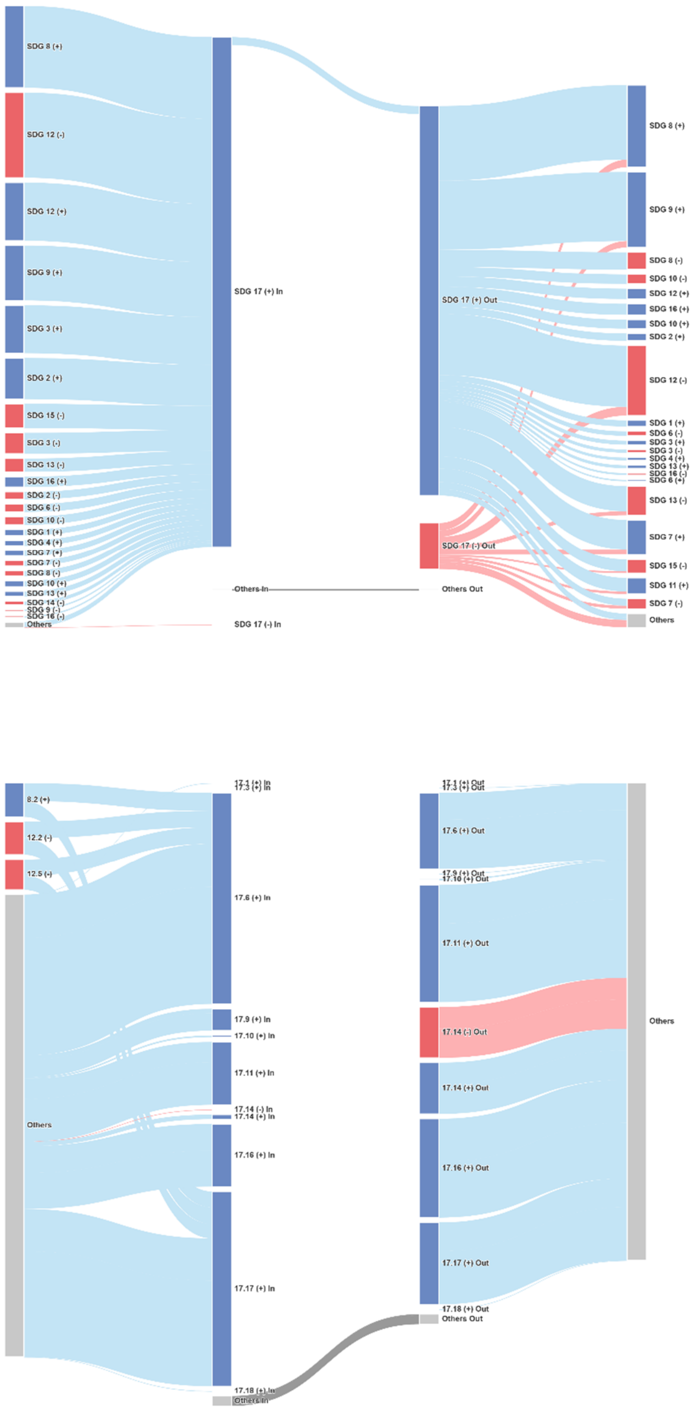

Fig. S19. SDG 17 (Partnerships for the Goals)

Table S1. Cross-model directional agreement and structural regime similarity

Model A	Model B	Directional agreement	ARI
GPT-5.2	Gemini 3.0 Pro	0.918	0.734
GPT-5.2	Claude 4.5 Opus	0.858	0.603
Gemini 3.0 Pro	Claude 4.5 Opus	0.864	0.702

Table S2. Validation of industry–SDG mappings against UNEP-FI sector-impact ground truth

Alignment regime	Precision
Classification alignment	0.700
Assumption-derived alignment	0.793
Semantic alignment	0.780

Table S3. R² comparison across functional forms

Domain	Current	Linear	Logarithmic	Polynomial (deg2)	Piecewise Linear
Engine	0.7373	0.4954	0.7359	0.6722	0.7053
Governance	0.7095	0.3048	0.6390	0.4822	0.6876
Environment-A	0.1175	0.0888	0.1157	0.1127	0.1286
Environment-B	0.1814	0.1477	0.1817	0.1681	0.1857
Transition	0.2569	0.1474	0.2558	0.2714	0.2792

Table S4. 10-year-ahead hindcast performance of parameter extrapolation models

Regime	**MAPE ARIMA(0,1,0)**	**MAPE ARIMA (others, mean)**	**MAPE Linear trend**	**80% Coverage ARIMA(0,1,0)**	**95% Coverage ARIMA(0,1,0)**	**80% Coverage ARIMA (others)**	**95% Coverage ARIMA (others)**
Engine	0.094	0.097	0.100	0.082	0.114	0.069	0.098
Governance	0.106	0.110	0.109	0.086	0.119	0.080	0.119
Environment	0.372	0.368	0.369	0.212	0.369	0.237	0.373
Environment-A	0.191	0.159	0.173	0.061	0.264	0.201	0.416
Environment-B	0.455	0.465	0.460	0.281	0.416	0.253	0.353
Transition	0.077	0.078	0.075	0.224	0.335	0.505	0.584