

A Global Task-Skill Atlas of Automation Exposure

Prashant Garg
Imperial College London

Jasmin Baier
University of Oxford

Tommaso Crosta
Bocconi University

March 15, 2026

Abstract

Automation exposure is often measured as if it were a context-free property of occupations or tasks, even though automation arrives through concrete work activities, depends on complementary skills, and varies across countries. We build a global task-skill atlas of automation exposure designed around that problem. The atlas combines validated many-to-many links between tasks and skills with a country-conditioned automation layer and a product/service-task bridge to technology-bearing goods and services.

The atlas links 6,275 ESCO skills to 15,586 O*NET tasks through 68,824 retained edges. We then classify the full 23,851-task universe under a common four-part measurement framework covering exposure level, technology channel, labor margin, and implementation dependency, and apply that framework to 142 countries. This yields 3,386,842 country-task labels with zero residual normalization failures and supports a GDP-weighted synthetic global benchmark built from country-conditioned task labels, which we compare directly to a matched context-free benchmark.

That replacement changes the measured global benchmark materially. Across tasks, channel concordance between the matched context-free benchmark and the GDP-weighted synthetic global is 0.788, while margin concordance is 0.860 and dependency concordance is 0.855. The synthetic global is nonetheless robust to weighting choices, with task-level Spearman concordance between 0.970 and 0.988 across GDP-weighted, population-weighted, and unweighted country means. Cross-country heterogeneity is also substantive: average task exposure ranges from 0.315 to 2.101 across countries, and tasks with larger cross-country dispersion also tend to move more when the benchmark becomes country-conditioned.

We further connect the atlas to 23,827 retained product-task links, of which 17,212 carry explicit role labels, and to goods-side country-task projections covering 11.8 million rows across 141 goods-trade reporters. The contribution is therefore not a single index but a reusable work-technology measurement system linking tasks, skills, countries, products, and trade-facing technology channels. The Automation Atlas measures feasibility, mechanism, and implementation constraints; it does not estimate realized adoption, displacement, wages, productivity, or causal labor-market effects.

1 Introduction

Automation exposure is still often measured as if it were a stable property of occupations or a context-free property of tasks. That is convenient for benchmarking, but it is not how technological change enters work. Automation arrives through specific work activities, depends on complementary skills, and is filtered by country-specific deployment conditions. When measurement collapses those layers too early, it becomes difficult to distinguish technical feasibility from implementation friction, substitution from augmentation, or global averages from the country-level variation that

gives them meaning (Autor et al., 2003; Acemoglu and Restrepo, 2018; Felten et al., 2021; Webb, 2020; Eloundou et al., 2023).

This paper builds a global task–skill atlas of automation exposure designed around that problem. The atlas starts from validated many-to-many links between tasks and skills, extends those links with a country-conditioned automation layer executed for 142 countries under a common measurement framework, and then links the same classification system to technology-bearing goods and services. The main empirical consequence is that the global benchmark itself changes. We construct a GDP-weighted synthetic global benchmark from country-conditioned task labels and compare it directly to a matched context-free benchmark built under the same framework without country context.

The argument sits at the intersection of three literatures. The first is task-based economics, which has long emphasized that technological change operates through the content of work rather than only through job titles (Autor et al., 2003; Acemoglu and Restrepo, 2018, 2019). The second is the growing measurement literature on automation and AI exposure, much of which has produced influential benchmarks but has often compressed exposure into occupation-level scores, single technology channels, or scalar global measures (Felten et al., 2021; Frey and Osborne, 2017; Webb, 2020; Eloundou et al., 2023, 2024). The third is a broader network and capability tradition, including skill-network and economic-complexity style work, that treats relational structure as part of the economic phenomenon rather than as a preprocessing detail (????).

Our contribution is to bring those strands together in one measurement architecture. First, we preserve many-to-many task–skill links rather than forcing thin crosswalks, so the structure of work remains part of what is being measured. Second, we show that country conditioning materially changes the global benchmark and that cross-country heterogeneity is a result, not a robustness footnote. Third, we add product and trade-facing layers that make the atlas operationally interpretable through observable channels of technology transmission. This extension connects the task–skill and country results to technology-bearing goods and services.

Throughout, the scope boundary remains explicit. The atlas measures technical feasibility, mechanism, and implementation constraints. It does not estimate realized adoption, worker displacement, wages, productivity, or causal labor-market outcomes. Those outcomes depend on organizational redesign, complementary capital, regulation, and policy choices that sit outside the present paper’s measurement objective. The atlas is designed instead as a transparent empirical foundation that later causal or institutional work can attach to.

The rest of the paper follows the same logic. Section 2 defines the data layers and reporting levels. Section 3 explains the measurement architecture, including task–skill construction, the task-level automation framework, country-conditioned aggregation, and product-role linkage. Section 4 presents the results in four blocks: task–skill structure, replacement of the context-free benchmark with a country-conditioned synthetic global benchmark, cross-country heterogeneity, and the product/trade-facing analysis. The Discussion then turns from measurement to interpretation, limitations, and the kinds of empirical work the atlas now makes possible.

2 Data and Measurement Layers

2.1 Task–skill structure

The task–skill layer combines ESCO skills with O*NET tasks. The validated structure retains 68,824 skill–task edges connecting 6,275 skills to 15,586 tasks. These retained edges are the reusable substrate for all downstream layers, including country-conditioned automation summaries, external benchmark harmonization, and product/service linkage.

2.2 Automation labeling universe

The automation layer is broader than the retained backbone because it labels the full O*NET task universe rather than only the tasks that survive skill-link pruning. The automation universe spans 23,851 standardized tasks. Each task receives a multidimensional label: exposure level, dominant channel, labor margin, and dominant implementation dependency, while preserving the underlying run-level record required for majority-vote aggregation and later diagnostics.

2.3 Country-conditioned layer

The country-conditioned automation layer applies the same task framework for 142 countries. The final root contains 3,386,842 country-task rows with zero normalization failures and full country coverage at task, occupation, industry, and skill levels. Those country-level labels are used directly for heterogeneity results and indirectly to build the GDP-weighted synthetic global benchmark that replaces the matched context-free benchmark in the main text. For descriptive region-income summaries, the paper uses a conventional seven-region split, so Pakistan and Afghanistan are grouped with South Asia rather than with the post-2025 World Bank MENAAP classification.

2.4 Product and trade-facing layer

The product and service layer uses curated HS and CPC references to connect technology-bearing goods and services to the task classification system. It retains 23,827 product-task edges and classifies 17,212 of them into explicit role labels. On the trade-facing side, the goods branch yields 11,840,489 country-task rows and 12,440,576 country-task-role rows across 141 goods-trade reporters. Services remain partial but usable as a secondary Supplementary Materials layer, with 37 reporters in the available broad-services source.

Object	Identifier	Count	Role
Retained skill-task backbone	skill \times task	68,824	Structural backbone
Retained skills	ESCO skill ID	6,275	Skill projection objects
Retained tasks	O*NET task code	15,586	Task-side backbone nodes
Same-schema context-free task universe	O*NET task code	23,851	Context-free benchmark
Country-conditioned task labels	country \times task	3,386,842	Country layer
Product-task retained edges	product \times task	23,827	Product linkage module
Role-ready product-task edges	product \times task	17,212	Role composition module
Goods-side country-task projections	country \times task	11,840,489	Trade-facing module

Table 1: Canonical measurement objects in the no-legacy paper stack.

3 Measurement Architecture

The measurement architecture is designed to keep the target of measurement explicit at each stage. The task-skill layer defines which skills and tasks are linked. The task-classification layer defines which automation pattern is being measured. The country-conditioned layer changes the benchmark from a context-free label to an aggregation over country-specific task labels. The product/service layer then connects those measures to observable technology-bearing goods and services. The guiding rule throughout is to preserve heterogeneity until aggregation is analytically necessary, and to make every aggregation step inspectable.

3.1 Many-to-many task–skill construction

The backbone is built through a recall-first, precision-controlled pipeline. Candidate links between skills and tasks are generated semantically, mapped deterministically to O*NET task identifiers through embedding retrieval, and then pruned through repeated validation. Retained links are those that survive majority voting under a conservative rule. This means that recall is handled upstream and precision is imposed downstream, rather than asking a single pass to do both. The result is a reusable many-to-many network rather than a thin crosswalk.

3.2 Task-level automation framework

Each task is labeled along four dimensions: exposure level, dominant technology channel, labor margin, and dominant implementation dependency. The framework is intentionally multidimensional. A scalar exposure score alone cannot distinguish, for example, between an LLM-mediated augmentation task and a robotics-heavy substitution task with large integration requirements. The task-level measure is therefore not simply “how exposed is this task,” but also “through what mechanism,” “with what likely adjustment margin,” and “subject to which deployment bottleneck.”

3.3 Country-conditioned automation and synthetic global replacement

Let E_{ct} denote the country-conditioned exposure of task t in country c , and let w_c denote GDP weights. The headline global benchmark used in this paper is

$$\bar{E}_t^{GDP} = \sum_c w_c E_{ct}, \quad \sum_c w_c = 1.$$

We report this GDP-weighted synthetic global benchmark together with cross-country dispersion and use a matched context-free benchmark as the direct comparator. Because both benchmarks use the same label set, structured response format, and aggregation logic, with country/year context removed only from the comparator, the comparison stays economically interpretable: what changes is the presence or absence of country context, not the classification framework itself.

3.4 Prompt contract, aggregation, and safeguards

The task-labeling pipeline uses a structured response format rather than unconstrained free text. Each run returns an explicit label set over exposure level, dominant channel, labor margin, and implementation dependency. Raw outputs are normalized, tied fields are recorded, and country- or context-specific task labels are aggregated by majority rule at the task level. The final country-conditioned layer has zero residual normalization failures, and the matched context-free run uses the same label set, aggregation logic, and normalization layer with country/year context removed. These safeguards turn what could otherwise look like opaque model output into an auditable measurement pipeline with explicit failure handling and conservative aggregation rules.

These safeguards matter for interpretation. They make the context-free and country-conditioned benchmarks directly comparable, and they create a transparent place to handle edge cases such as empty outputs, normalization failures, or safety-triggered title overrides. The one material safety-triggered task in the workflow was resolved through a semantically equivalent paraphrase policy that preserved task meaning while avoiding blocked wording. Throughout, the paper reports exposure and mechanism measures, not realized adoption or labor-market outcomes.

3.5 Product/task linkage and role classification

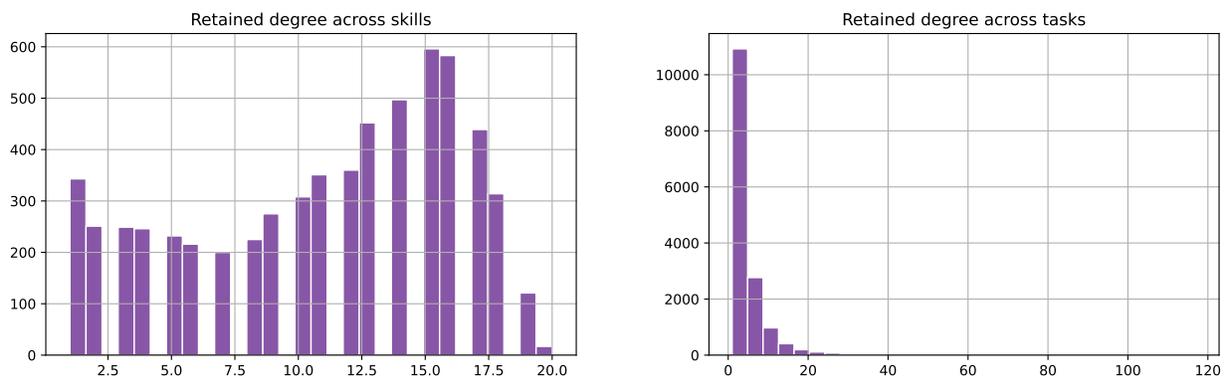
The product branch links HS/CPC items to tasks through title normalization, productive-universe curation, product-to-task matching, edge pruning, and role classification. Role labels distinguish technology arriving as autonomous execution, machine execution under supervision, or human use of a tool or machine. This turns abstract task labels into technology-bearing goods and service interfaces and creates a direct bridge to trade-facing country-task diagnostics. The goods branch uses a locally archived BACI 202601 dataset, while the tariff Supplementary Materials layer uses WITS/TRAINS.

4 Results

The results answer four linked questions. What changes when many-to-many links between tasks and skills are preserved rather than compressed? What changes when a matched context-free benchmark is replaced with a country-conditioned synthetic global benchmark? How large is the cross-country heterogeneity that motivates that replacement? And how do those measures connect to observable channels of technology transmission through products and trade?

4.1 Backbone structure

The first question is whether preserving many-to-many links between tasks and skills actually changes measured exposure or merely adds presentational complexity. The retained network remains the structural core of the system. It contains 68,824 edges connecting 6,275 skills to 15,586 tasks, with mean skill degree 10.97 and mean task degree 4.42. Those averages already show why thin one-to-one mappings are inadequate: typical skills connect to multiple tasks and typical retained tasks connect to multiple skills. The network therefore carries combinatorial information that would be difficult to preserve under early compression.



(a) Skill degree distribution in the retained backbone.

(b) Task degree distribution in the retained backbone.

Figure 1: Degree distributions for retained skills and tasks in the validated backbone. Multi-linking is common on both sides of the network, which is why thin crosswalks discard relational structure that may later matter for measurement; the figure describes connectivity, not realized labor-market impact.

The consequence is direct rather than merely structural. Figure 2 compares skill exposure ranks under the preserved backbone to ranks obtained after forcing every task to keep only a single skill. Under the matched context-free benchmark, 194 skills enter the network top decile and 218

leave the one-to-one top decile, which means that 43.9% of the network top decile is network-only. Under the GDP-weighted synthetic global, the corresponding figure is 39.6%. Mean absolute percentile-rank gaps remain large under both benchmarks, at 0.136 and 0.118 respectively, across 4,537 comparable skills. Supplementary Tables S1–S3 and Supplementary Figures S1–S2 show that these shifts are neither trivial nor confined to a tiny fringe of the skill space. Supplementary Table S4 and Supplementary Figure S3 show that the same compression-loss logic survives downstream when the preserved backbone is projected into occupation- and industry-facing summaries.

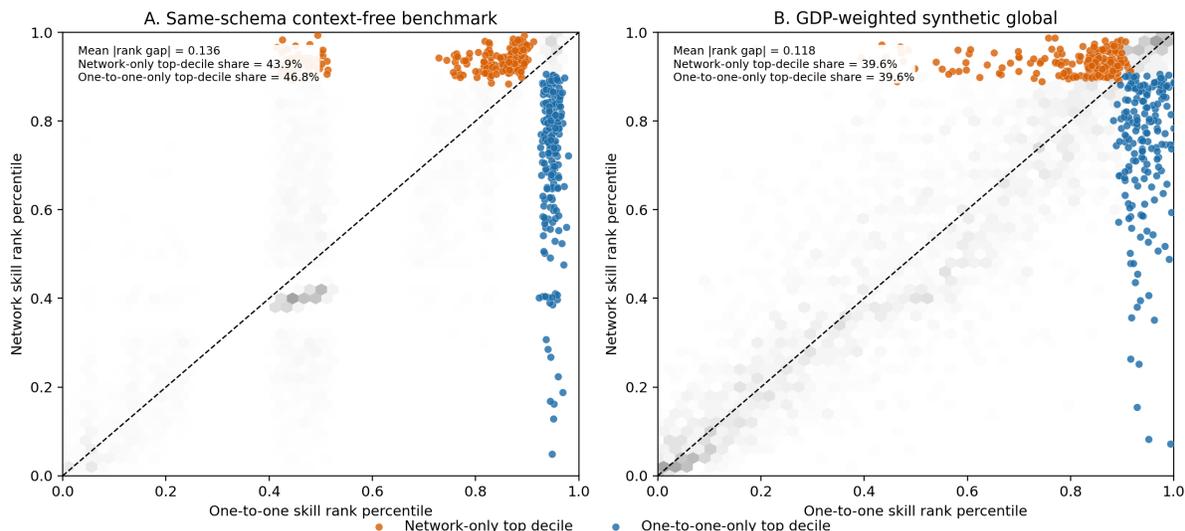


Figure 2: Direct reranking consequence of preserving the many-to-many backbone rather than forcing each task to keep only one skill. Orange points are skills that enter the network top decile only when the full backbone is preserved; blue points are skills that appear in the one-to-one top decile only under the compressed map. The effect remains material under both the matched context-free benchmark and the GDP-weighted synthetic global.

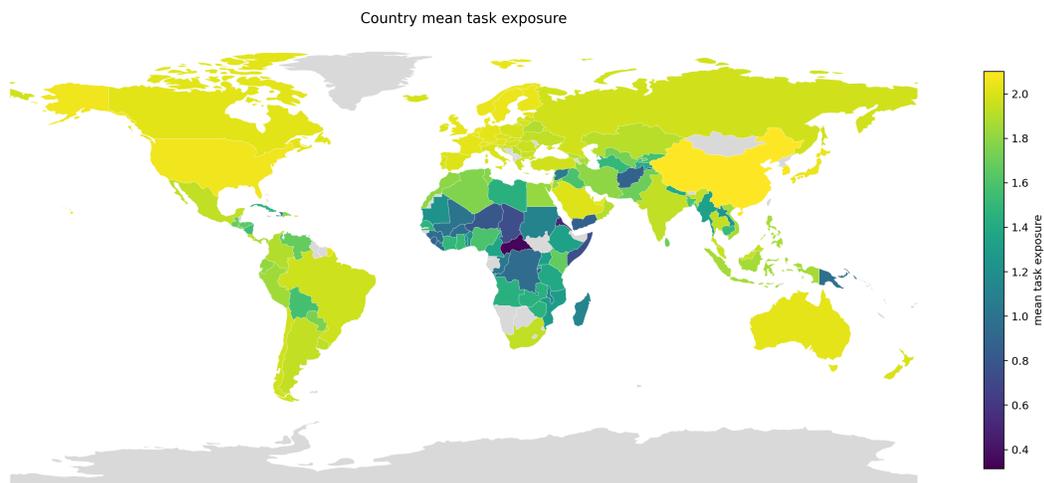
This structure is therefore not a presentational detail. It is the mechanism by which task-side automation diagnostics reach skills, occupations, industries, and later product-linked trade measures. The backbone is not merely an input table but the organizing structure of the paper, and the reranking exercise shows that collapsing it early would already change which skills appear most exposed before any country, product, or trade layer is added.

4.2 From a matched context-free benchmark to a country-conditioned synthetic global

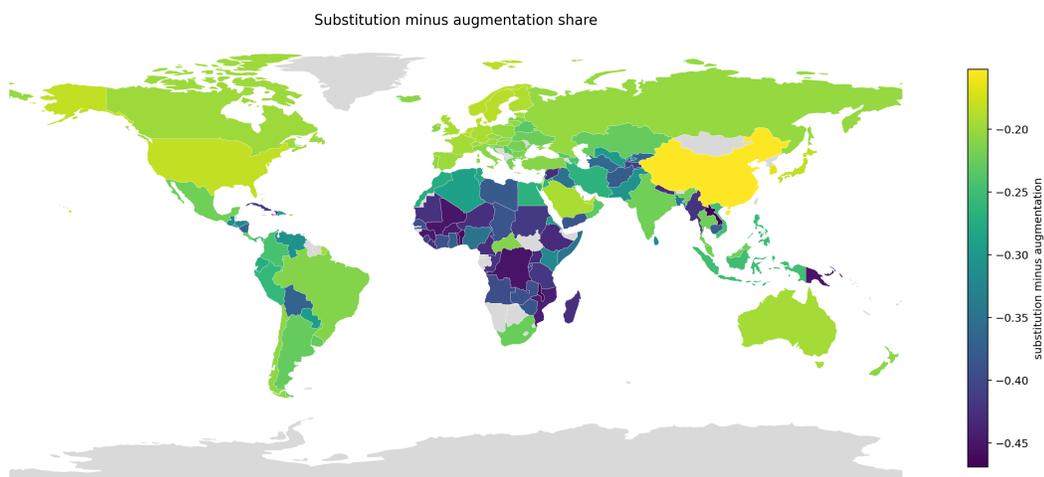
The second question is whether country conditioning changes the measured global benchmark in a substantively meaningful way. The main reported global value is the GDP-weighted synthetic global constructed from the 142-country layer. To make that choice auditable, we compare a matched context-free benchmark directly to the GDP-weighted synthetic global.

Across 23,851 tasks, task-level channel concordance between the matched context-free benchmark and the GDP-weighted synthetic global is 0.788. Margin concordance is 0.860 and dependency concordance is 0.855. Mean absolute rank change is 6,037 task positions and 49.2 occupation positions. The shift is therefore large enough to matter for substantive interpretation, not merely for presentation. The reason is not that the classification framework changes across runs. It is that the benchmark itself changes once country context is allowed to shape task exposure and the global

differences can be related to country covariates in a disciplined descriptive way. The weighting-and-dispersion panel then shows that the headline synthetic-global result is not fragile to alternative country aggregation schemes and is largest precisely where country dispersion is greatest.



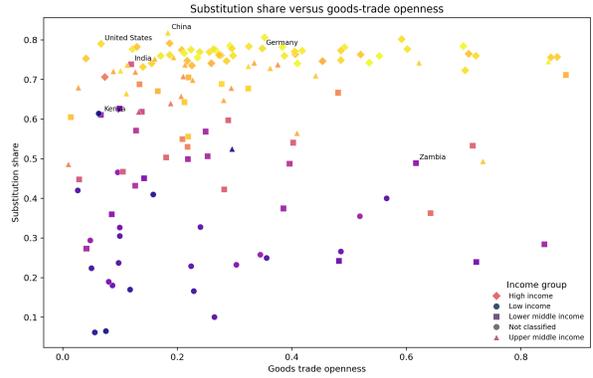
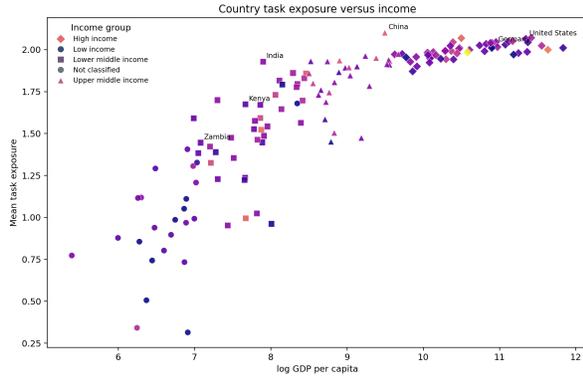
(a) Mean task exposure across the 142-country layer.



(b) Substitution share minus augmentation share.

Figure 4: Country-level maps of mean task exposure and substitution-minus-augmentation orientation across the 142-country layer. The point is dispersion in both level and adjustment direction, not a causal claim about geography or institutions by themselves.

Task-level heterogeneity is especially informative because it shows where global means hide local context sensitivity. Some tasks remain tightly clustered around their global mean. Others with similar average exposure exhibit much wider cross-country spread, implying that any single global point estimate should be interpreted with caution unless dispersion is reported alongside it. This is also why the synthetic-global replacement is not a cosmetic relabeling: tasks with higher country dispersion are disproportionately likely to move when the benchmark becomes country-conditioned. The same propagation is visible at the skill level. Supplementary Table S12 and Supplementary



(a) Mean exposure and development level.

(b) Substitution share and goods-trade openness.

Figure 5: Country-level benchmark scatters linking atlas outputs to development and trade covariates. These panels show that the country layer is suitable for comparative benchmarking, while remaining descriptive rather than causal.

Region	Income group	Countries	Mean exposure	Substitution share	Goods trade intensity
East Asia & Pacific	High income	7	2.03	0.78	2.15
East Asia & Pacific	Lower middle income	6	1.46	0.50	1.63
East Asia & Pacific	Upper middle income	4	1.96	0.75	2.12
Europe & Central Asia	High income	31	2.00	0.76	2.07
Europe & Central Asia	Lower middle income	3	1.58	0.56	1.53
Europe & Central Asia	Upper middle income	8	1.86	0.70	1.86
Latin America & Caribbean	High income	5	1.94	0.74	1.90
Latin America & Caribbean	Lower middle income	4	1.44	0.49	1.32
Latin America & Caribbean	Not classified	1	1.68	0.61	1.20
Latin America & Caribbean	Upper middle income	11	1.80	0.67	1.72
Middle East & North Africa	High income	7	1.98	0.76	2.00
Middle East & North Africa	Low income	2	0.95	0.26	1.37
Middle East & North Africa	Lower middle income	6	1.78	0.67	1.63
Middle East & North Africa	Upper middle income	4	1.65	0.60	1.63
North America	High income	2	2.04	0.78	2.10
South Asia	Low income	1	0.88	0.24	0.84
South Asia	Lower middle income	5	1.66	0.60	1.63
Sub-Saharan Africa	Low income	19	0.92	0.25	1.01
Sub-Saharan Africa	Lower middle income	14	1.38	0.45	1.33
Sub-Saharan Africa	Not classified	1	1.33	0.42	1.23
Sub-Saharan Africa	Upper middle income	1	1.93	0.74	1.86

Table 3: Country summary by conventional seven-region split and income group. Pakistan and Afghanistan are grouped in South Asia rather than in the FY26 World Bank MENAAP grouping.

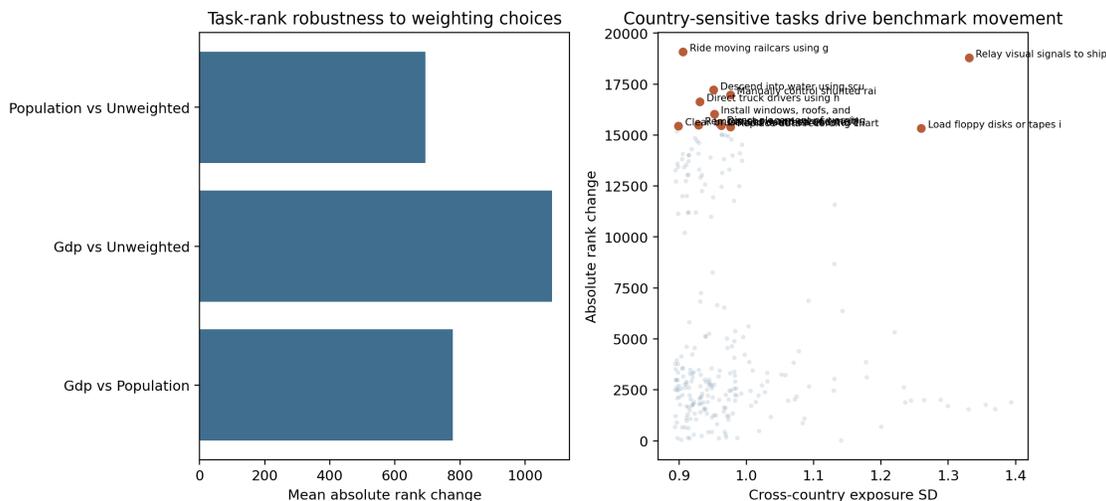


Figure 6: Weighting and dispersion robustness for the country-conditioned synthetic global. The main benchmark remains stable across GDP, population, and unweighted country means, while larger context-free-versus-synthetic shifts are concentrated in more dispersion-prone tasks.

Figure S11 show that some backbone-projected skills exhibit much wider cross-country spread than their average exposure alone would suggest. Heterogeneity is therefore the proof that the benchmark correction matters, not an auxiliary descriptive add-on.

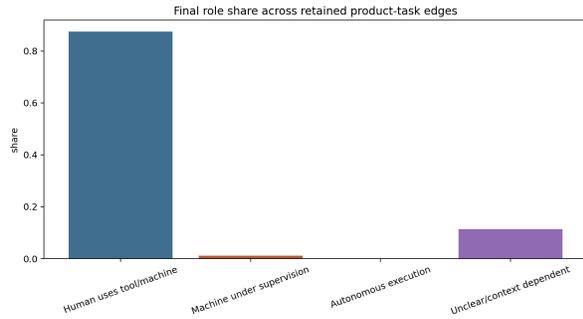
This is also the point where the product/trade analysis becomes useful rather than decorative. Once we know that countries differ materially in exposure, channel mix, and adjustment orientation, the next question is how those differences might connect to observable channels of technology transmission.

4.4 Product and trade-facing analysis

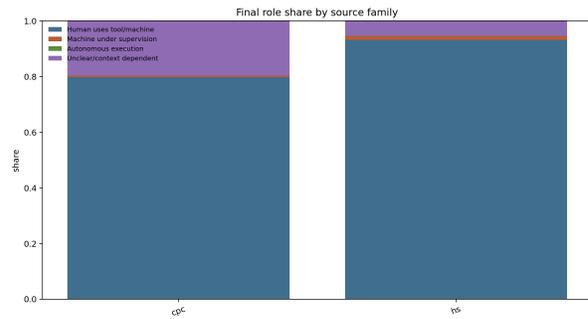
The fourth question is how these measures connect to observable channels of technology transmission. The product and service layer is now a substantive empirical analysis rather than an illustrative add-on. On the product side, 23,827 retained product-task edges survive pruning and 17,212 of those carry explicit role labels. That subset is large enough to support descriptive composition analysis and goods-side trade-facing task projection. Using a locally archived BACI 202601 dataset leaves the central message unchanged while providing clear data provenance for the goods-side results.

Figure 7 shows how technology-bearing goods and services enter the atlas under different role labels. Figures 8 and 9 then trace those differences at country level. These panels are descriptive mechanism evidence rather than alternative exposure measures: countries with similar exposure can still differ in the observable goods-side channels, role mixes, and trade-facing technology interfaces through which automation may arrive. Supplementary Figure S16 shows that the core goods-side pattern remains visible after trimming the single most extreme machine-execution outlier for display.

The goods branch projects those product-task interfaces into 11.8 million country-task rows covering 141 goods-trade reporters. Services remain partially covered and therefore stay out of the main text, but they are sufficiently developed to support Supplementary Materials diagnostics. The broader point is conceptual: the atlas is not only a task–skill atlas. It is a measurement system that links tasks, skills, countries, products, and trade-facing channels, and that makes the country-



(a) Overall role composition of finalized product-task edges.



(b) Role composition by source family.

Figure 7: Role composition of finalized product-task edges overall and by source family. The product layer distinguishes whether technology-bearing goods and services enter work as autonomous execution, supervised machine execution, or human-use tools; it does not, by itself, identify realized adoption.

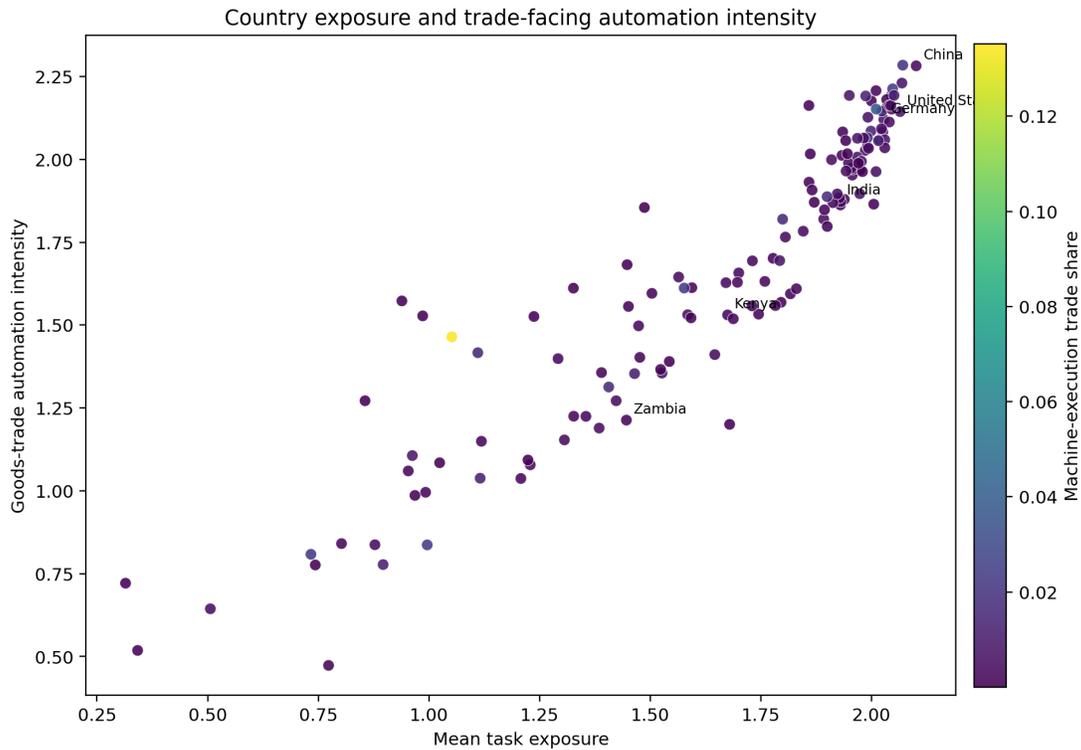


Figure 8: Country-level relationship between atlas exposure and trade-facing automation intensity. The vertical axis records the intensity of projected goods-side automation channels, while point color records the share of goods trade value arriving through machine-execution roles. Countries with similar exposure can therefore differ both in overall trade-facing automation intensity and in the role mix through which technology-bearing goods may arrive.

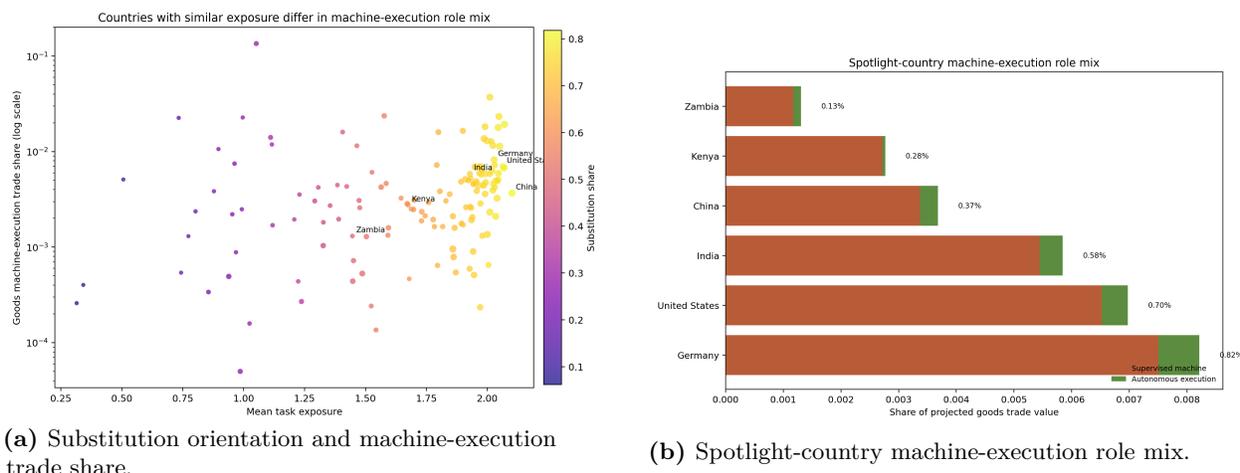


Figure 9: Trade-facing refinement panels for substitution orientation, machine-execution trade share, and spotlight-country role mixes. The left panel shows that countries with similar average exposure can still differ sharply in the goods-side role mix through which technology-bearing products arrive. The right panel shows that even among spotlight countries, machine-execution channels remain quantitatively modest but vary enough to refine the country-conditioned result with observable transmission detail.

Metric	Value
Retained product-task edges	23,827
Role-ready product-task edges	17,212
Unresolved retained edges	6,615
Goods-side country-task rows	11,840,489
Goods-trade reporters	141
Services-side country-task rows	3,067,551

Table 4: Coverage and readiness of the refreshed product linkage branch.

Country	ISO3	Goods intensity	Auto-exec share	Exposed goods trade pc
Singapore	SGP	2.28	0.000	155,514
China	CHN	2.28	0.000	7,328
Korea, Rep.	KOR	2.23	0.001	37,176
Netherlands	NLD	2.21	0.000	64,286
Hong Kong SAR, China	HKG	2.21	0.000	28,647
Denmark	DNK	2.19	0.001	39,921
Malaysia	MYS	2.19	0.000	31,786
Japan	JPN	2.19	0.001	20,309
Iceland	ISL	2.19	0.000	9,040
United Arab Emirates	ARE	2.18	0.000	29,465
Ireland	IRL	2.18	0.000	56,525
Viet Nam	VNM	2.16	0.000	11,423

Table 5: Illustrative country-level trade-facing automation measures.

conditioned benchmark more interpretable by showing one set of observable pathways through which automation technologies may reach different economies.

That extension matters because it keeps the paper from ending at benchmark replacement alone. The atlas does not merely say that the global benchmark should change; it also shows how the corrected exposure measure can be connected to concrete technology-bearing interfaces that vary across countries.

5 Discussion

The paper changes the interpretation of automation exposure in three related ways.

First, it shows that preserving many-to-many links between tasks and skills is not a technical embellishment. If automation exposure operates through tasks and the skills attached to them, then early compression removes part of the structure of work before interpretation even begins. The direct consequence exercise shows that forcing one skill per task changes roughly two-fifths of the network top decile under both reported benchmarks. Those links are therefore not just infrastructure; they are what allow exposure to propagate meaningfully across skills, occupations, industries, and later product-linked trade channels.

Second, it changes what should count as the main global benchmark. A matched context-free benchmark remains useful as a disciplined comparator, but it is no longer the best headline measure once country-conditioned labels are available at scale. The GDP-weighted synthetic global is the more appropriate headline benchmark because it is built from the country-conditioned task layer rather than assumed independently of it. The cross-country results show why this matters: global means can hide substantial variation in exposure level, adjustment orientation, and implementation bottlenecks.

Third, it shows that automation exposure becomes more interpretable when it is linked to technology-bearing goods and services. The product/trade analysis does not convert the paper into a causal trade study, but it does add a concrete transmission layer. Countries with similar task exposure can still differ in the goods-side role composition and trade-facing technology channels through which automation may arrive. That helps connect the atlas to observable economic interfaces rather than leaving it at the level of abstract task labels alone.

These gains come with clear boundaries. The paper does not claim realized adoption. It does not claim worker displacement. It does not estimate wages, productivity, or equilibrium incidence. Country conditioning remains model-mediated rather than observed institutional realization. Services coverage is partial, tariff context is available only for a subset of reporters, and the trade-facing evidence is descriptive rather than causal. The atlas is a structured diagnostic of feasibility, mechanism, and implementation constraints. That boundary is a strength rather than a weakness: it keeps the present contribution focused on measurement while making it easier for later causal, institutional, or policy work to build on a transparent empirical base.

6 Conclusion

This paper builds a global task–skill atlas of automation exposure and then shows why that atlas must be extended rather than compressed. Many-to-many links between tasks and skills remain the structural core. The headline global benchmark is not treated as context-free, but rebuilt as a GDP-weighted synthetic global benchmark from a completed 142-country task layer. A product and service layer then links the same classification system to technology-bearing goods and services and makes the resulting exposure measures more operationally interpretable.

The central implication is simple: global automation exposure should be measured as an aggregation of country-conditioned task exposure, not as a single context-free benchmark. Preserving task–skill links shows how exposure propagates through the structure of work; adding the country layer shows why the benchmark itself changes; and adding the product/trade layer shows one set of observable channels through which those measured exposures may matter. The result is not a forecast of realized labor-market outcomes, but a reusable empirical system for studying work, automation, and technology transmission under a common classification system.

Supplementary Materials

Supplementary Note A Data Objects and Construction Details

The Supplementary Materials consolidate the supporting diagnostics for the two benchmark measures used in this paper: the matched context-free benchmark and the 142-country GDP-weighted synthetic global. They also extend the main text with goods, services, and tariff-context materials that are not central enough to carry in the main results.

The goods results use a BACI 202601 archive. The tariff-context layer uses WITS/TRAINS and covers 76 reporters. The services branch still relies on the broad-services source used here rather than BaTIS, which is not included in this paper. Broader services and value-chain datasets such as BaTIS, OECD TiVA/ICIO, and BTDIxE remain natural extensions for later work.

The Supplementary Materials therefore document four things. First, the direct ranking consequences of preserving many-to-many task–skill links rather than compressing them into a one-to-one map. Second, how the external comparator suite lines up against the two benchmark measures used in this paper. Third, how weighting, dispersion, and country-specific task variation reinforce the main text result. Fourth, how trade, tariff, and services extensions broaden the empirical reach of the atlas.

Backbone consequence diagnostics

The main text shows that the retained network is densely multi-linked and that preserving it materially changes measured skill rankings. The diagnostics below make that consequence more explicit. Supplementary Table S1 summarizes the scale of reranking under both reported benchmarks, Supplementary Table S2 gives concrete entrant and displaced examples, and Supplementary Table S3 shows which broad skill families supply the largest number of network-only entrants. Supplementary Figures S1 and S2 then show where the largest residual shifts occur and how those entrants are distributed across families.

Benchmark	Comparable skills	Network-only top-decile entrants	Entrant share	One-to-one-only displaced	Mean abs rank gap
Same-schema context-free	4,537	194	43.9%	218	0.136
GDP-weighted synthetic global	4,537	180	39.6%	180	0.118

Table S1: Direct consequence of preserving the many-to-many backbone rather than forcing one skill per task.

Supplementary Note B External Benchmark Validation

The external comparator block shows how the two benchmark measures used in this paper line up against multiple benchmark families that encode different constructs, levels, and vintages. We retain Felten, Frey–Osborne, Webb, GPTs-are-GPTs, and Septiandri because each captures a different external notion of automation or AI relevance. The relevant question is not which benchmark “wins,” but whether the atlas measures behave coherently when compared to multiple established external references.

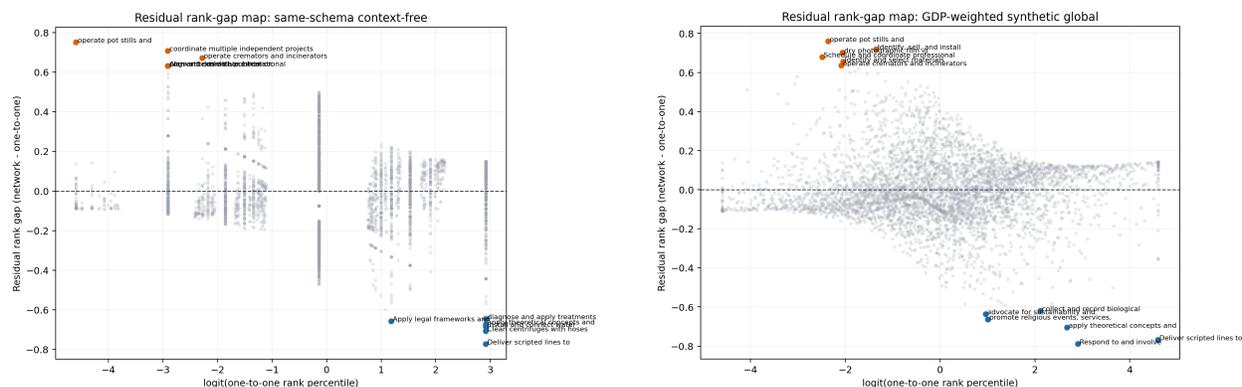
Two findings matter here. First, external agreement is moderate rather than perfect, which is expected because the benchmark families measure different constructs and vintages. Second, the matched context-free benchmark and the GDP-weighted synthetic global do not line up identically to those external families, which is exactly why it is useful to keep both in view: the paper needs a context-free comparator, but the headline global benchmark should still be country-conditioned.

Benchmark	Network-only top-decile entrants	One-to-one-only displaced skills
Same-schema context-free	manage and schedule patient appointments for audiologists Tend agitation machines for consistent batch mixing analyze and model data to support decision-making	Control vessel navigation, steering, and emergency safety promote and negotiate tourism services at industry events clean semiconductor wafers using automated equipment and c
GDP-weighted synthetic global	Review manager drafts for completeness, accuracy, and form remotely control ball mills for pulverising dried ingredie track credit commitments to detect anomalies and improper	Deliver scripted lines to bring a character to life mix multiple live video streams using specialized equipmen translate diverse text types across various professional d

Table S2: Illustrative skills that enter or exit the top decile when the preserved backbone replaces a one-to-one map.

Benchmark	Entrant family	Entrant skills
Same-schema context-free	skills	137
	transversal skills and competences	33
	knowledge	24
GDP-weighted synthetic global	skills	128
	transversal skills and competences	31
	knowledge	21

Table S3: Leading entrant skill families under the preserved backbone.



(a) Residual rank-gap map under the matched context-free benchmark.

(b) Residual rank-gap map under the GDP-weighted synthetic global.

Figure S1: Residual backbone-consequence diagnostics under the two benchmark constructions used in this paper. These panels highlight the skill labels with unusually large reranking residuals after conditioning on baseline one-to-one rank, making clear that the main-text consequence figure is not driven only by ties or visual crowding.

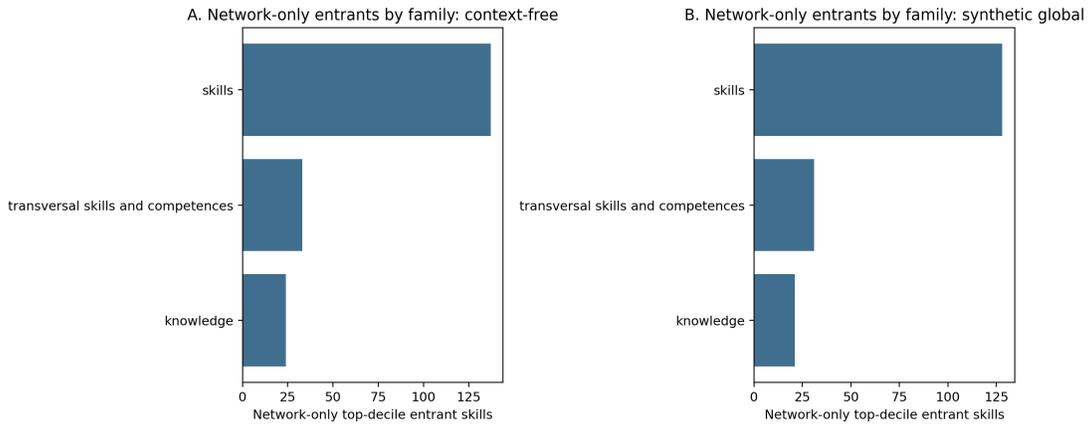


Figure S2: Leading entrant skill families under the preserved backbone. The family distribution shows that network-only entrants are not isolated curiosities; they cluster in interpretable groups that are repeatedly undervalued when tasks are forced to keep only one skill.

Benchmark	Level	Objects	Mean abs rank gap	Top-decile overlap
Same-schema context-free	Occupation	923	0.078	73.1%
Same-schema context-free	Industry	137	0.101	78.6%
GDP-weighted synthetic global	Occupation	923	0.074	77.4%
GDP-weighted synthetic global	Industry	137	0.091	71.4%

Table S4: Downstream occupation- and industry-side compression loss when the preserved backbone replaces a one-to-one map.

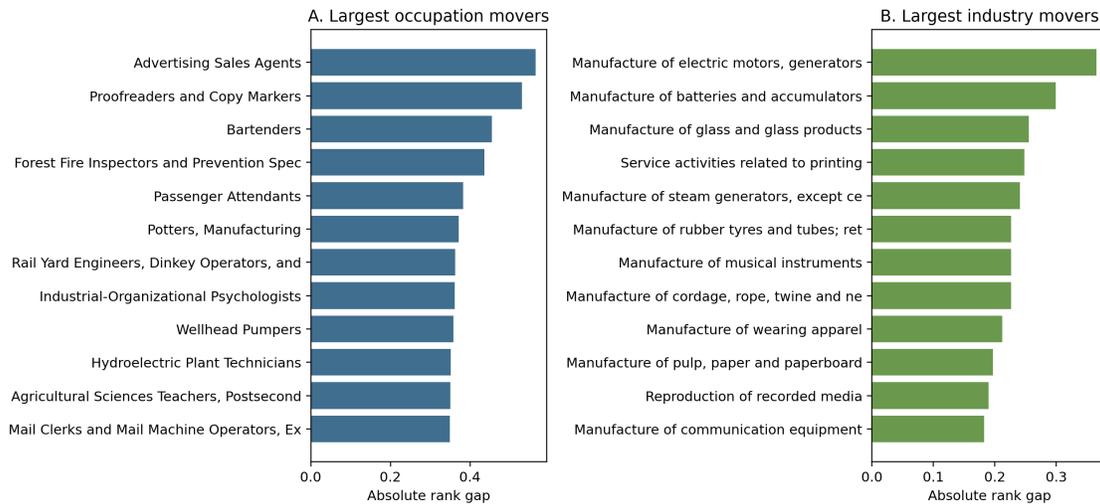


Figure S3: Downstream occupation- and industry-facing compression loss under the GDP-weighted synthetic global. The point is not to replace the main occupation or industry layers, but to show that the same backbone-compression logic survives when the preserved network is propagated into later reporting summaries.

Benchmark	Level	Comparable rows
Felten AIOE	occupation	665
Frey-Osborne	occupation	582
GPTs DV gamma	occupation	923
GPTs human gamma	occupation	923
Septiandri AII	occupation	639
Webb AI	occupation	762
Felten AIOE	task	17,111
Frey-Osborne	task	14,689
GPTs GPT-4 task	task	23,743
GPTs human task	task	116
Septiandri AII	task	17,904
Webb AI	task	19,671

Table S5: Coverage of external comparators in the no-legacy appendix benchmark suite.

Benchmark	Overlap	Pearson	Spearman	Mean abs rank change
Felten AIOE	17,111	0.156	0.171	5,134
Frey-Osborne	14,689	0.181	0.203	4,338
GPTs GPT-4 task	23,743	0.533	0.575	4,552
GPTs human task	116	0.764	0.832	15
Septiandri AII	17,904	0.127	0.143	5,438
Webb AI	19,671	0.083	0.088	6,202

Table S6: Task-level comparison of the GDP-weighted synthetic global benchmark against external task benchmarks.

Benchmark	Overlap	Pearson	Spearman	Mean abs rank change
Felten AIOE	665	0.313	0.261	190
Frey-Osborne	582	0.292	0.358	153
GPTs DV gamma	923	0.406	0.405	233
GPTs human gamma	923	0.387	0.359	241
Septiandri AII	639	0.221	0.271	175
Webb AI	762	0.138	0.178	228

Table S7: Occupation-level comparison of the GDP-weighted synthetic global benchmark against external occupation benchmarks.

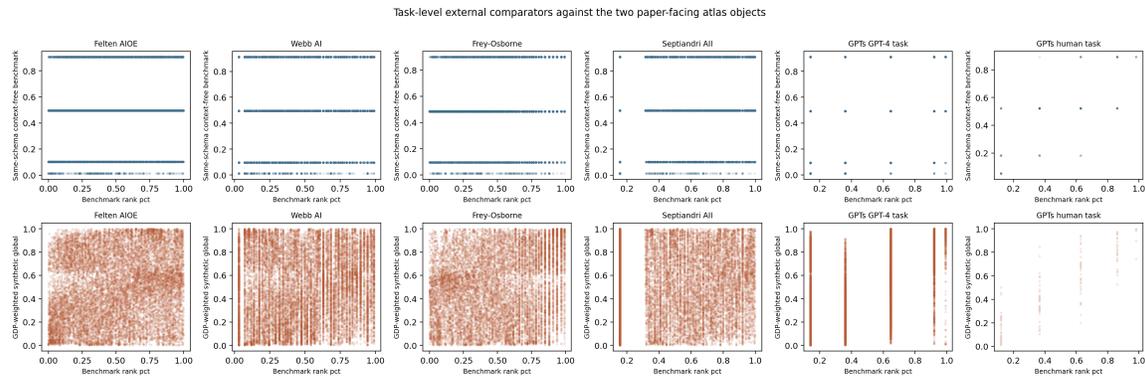


Figure S4: Task-level external comparator matrix across the retained validation suite. The purpose is triangulation rather than winner-take-all validation: moderate alignment is expected because the retained comparators differ in construct, unit, and vintage.

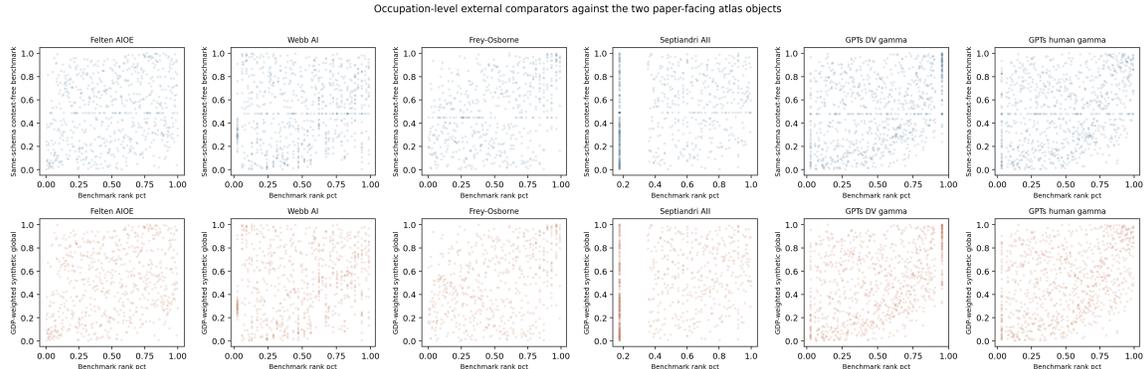


Figure S5: Occupation-level external comparator matrix across the retained validation suite. The panels show where agreement survives aggregation to occupations and where construct mismatch remains visible even after harmonization.

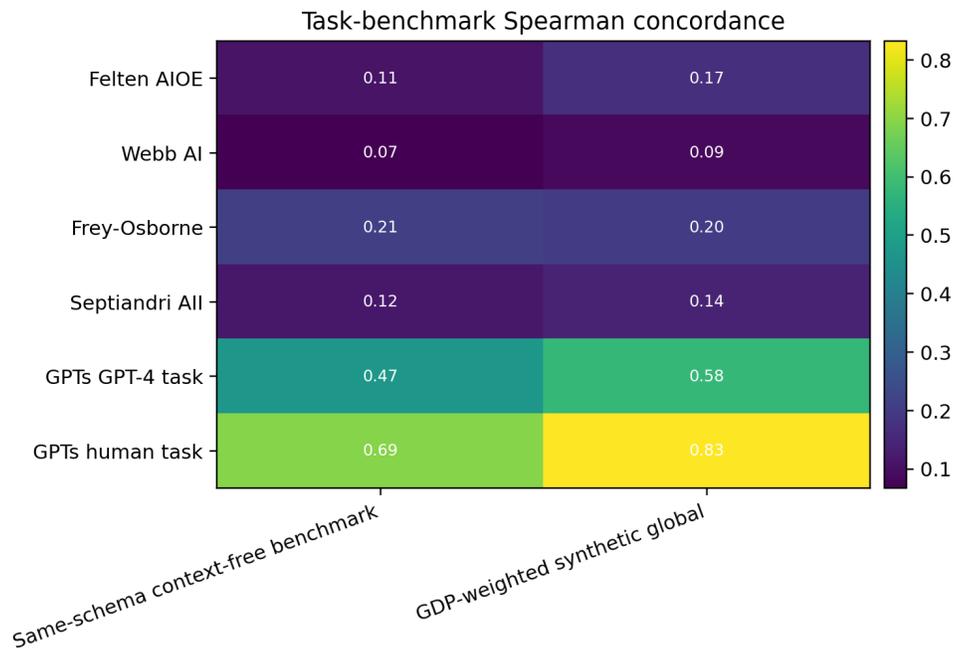


Figure S6: Correlation heatmap summarizing the external comparator suite on the harmonized benchmark measures used in this paper. The heatmap summarizes patterned agreement and disagreement across comparator families rather than collapsing them into a single validation score.

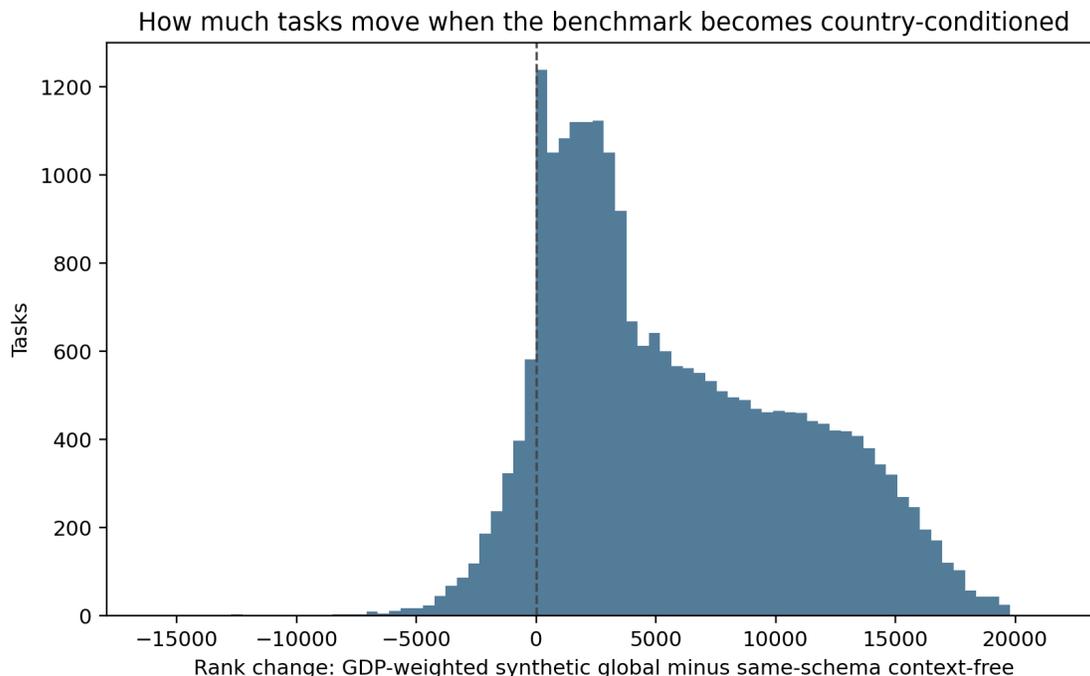


Figure S7: Rank-change distributions showing how tasks move when the benchmark changes from the matched context-free benchmark to the GDP-weighted synthetic global. These distributions connect the external-validation block back to the main text by showing that benchmark replacement has broad consequences rather than only a few isolated movers.

Supplementary Note C Country-Level and Trade/Tariff Robustness

The country layer now supports maps, benchmark scatters, and compact league tables. It also supports direct linkage to trade-facing measures. The goods branch now covers 141 reporters, and the tariff Supplementary Materials layer adds 76 reporters with matched MFN tariff context from WITS/TRAINS using the latest available year among 2023, 2022, and 2021. This section shows both that the main benchmark is robust to alternative country aggregation choices and that the country-conditioned atlas can already be joined to country- and border-level observables.

The tariff block is intentionally contextual rather than structural. It does not replace the BACI-based goods branch and it is not presented as a causal barrier estimate. Its role is narrower: to show that trade-facing exposure measures can be joined to policy-relevant border measures for a substantial subset of reporters. That makes it a suitable Supplementary Materials layer here, while richer services and value-chain layers remain future work.

Supplementary Note D Additional Heterogeneity and Product-Role Diagnostics

The remaining Supplementary Materials figures document the product-role branch more fully than is appropriate in the main text. They show how role votes are distributed, which products are structurally central in the linkage graph, and where goods-side HS2 trade linkage is most concentrated. These are descriptive diagnostics, but they also clarify why the product branch should

Level	Weighting pair	Spearman	Mean abs rank change
Industry	Gdp vs Population	0.983	5.3
Industry	Gdp vs Unweighted	0.966	7.5
Industry	Population vs Unweighted	0.994	3.1
Occupation	Gdp vs Population	0.978	40.6
Occupation	Gdp vs Unweighted	0.958	56.6
Occupation	Population vs Unweighted	0.994	21.3
Skill	Gdp vs Population	0.980	255.9
Skill	Gdp vs Unweighted	0.960	359.1
Skill	Population vs Unweighted	0.992	155.9
Task	Gdp vs Population	0.984	777.5
Task	Gdp vs Unweighted	0.970	1082.4
Task	Population vs Unweighted	0.988	693.6

Table S8: Robustness of the synthetic-global layer to alternative country weighting schemes.

Task	Exposure SD	Abs rank change	Abs exposure delta
Relay visual signals to ships using semaphores or blinker lights	1.331	18792	1.210
Ride moving railcars using grab irons and ladder steps	0.905	19084	1.321
Direct vehicle traffic using hand signals or signs.	0.887	20924	1.665
Descend into water using scuba gear or diving suits	0.951	17215	0.808
Manually control shunted railcar speed using handwheels while riding	0.976	16976	0.746
Pull railcar knuckles to open them for coupling	0.863	18922	1.182
Control tree fall direction using axes, chainsaws, and wedges	0.852	20677	1.604
Direct truck drivers using hand and horn signals for coverage	0.931	16648	0.672
Couple and uncouple air hoses and electrical connections between cars	0.875	17017	0.754
Set emergency flares and flags to warn oncoming trains	0.823	20585	1.585
Direct vehicle traffic and signal hazards using flags and signs	0.823	19478	1.331
Haul and spread soil conditioners using wheelbarrows and shovels	0.862	17106	0.776
Attend continuing education classes and professional development works	0.830	18225	1.009
Install windows, roofs, and trim onto vehicle bodies	0.952	16035	1.652
Send film to photofinishing laboratories for processing	0.811	19043	1.279

Table S9: Illustrative tasks that are simultaneously high-dispersion and high-shift.

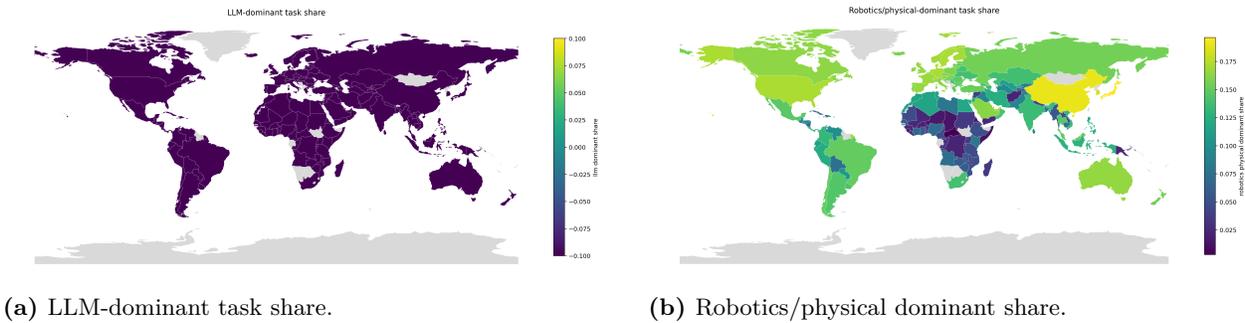


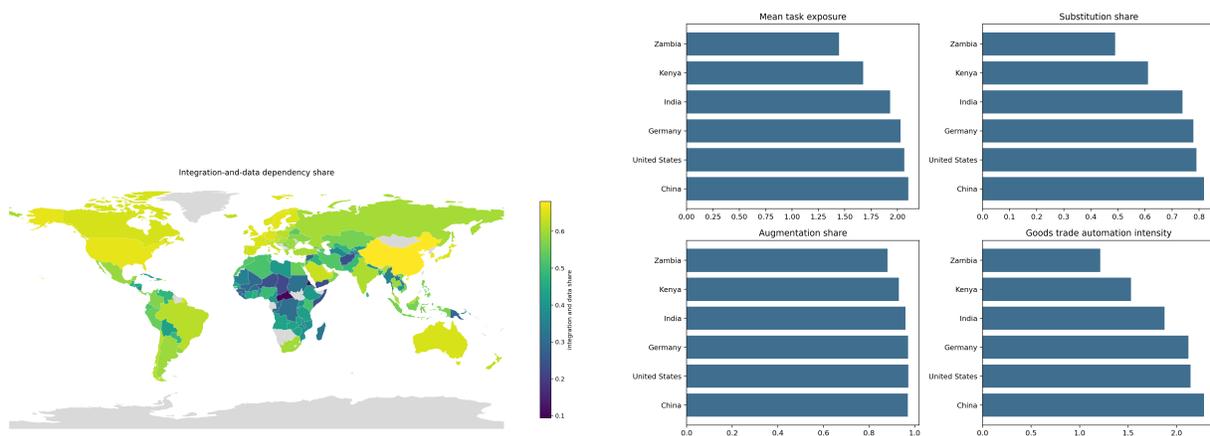
Figure S8: Supplementary country maps for dominant technology-channel composition. These maps extend the main text by showing that the country layer varies not only in exposure level but also in the dominant technological pathway associated with exposure.

Region	Income group	Countries	Mean exposure	Goods trade intensity
North America	High income	2	2.04	2.10
East Asia & Pacific	High income	7	2.03	2.15
Europe & Central Asia	High income	31	2.00	2.07
Middle East & North Africa	High income	7	1.98	2.00
East Asia & Pacific	Upper middle income	4	1.96	2.12
Latin America & Caribbean	High income	5	1.94	1.90
Sub-Saharan Africa	Upper middle income	1	1.93	1.86
Europe & Central Asia	Upper middle income	8	1.86	1.86
Latin America & Caribbean	Upper middle income	11	1.80	1.72
Middle East & North Africa	Lower middle income	6	1.78	1.63
Latin America & Caribbean	Not classified	1	1.68	1.20
South Asia	Lower middle income	5	1.66	1.63

Table S10: Highest-exposure region-income groups in the country layer, under the conventional seven-region split used in the paper tables.

Country	ISO3	Year	Mean MFN tariff	Share nonzero
Burundi	BDI	2023	27.68	0.97
Uganda	UGA	2023	26.87	0.94
Kenya	KEN	2023	25.15	0.81
Bangladesh	BGD	2023	24.43	1.00
Rwanda	RWA	2023	22.01	0.85
Tanzania	TZA	2023	21.20	0.81
India	IND	2023	19.87	0.99
Nepal	NPL	2023	19.25	1.00
Turkiye	TUR	2023	17.71	0.90
Ecuador	ECU	2023	17.47	0.89
Pakistan	PAK	2022	15.96	0.92
Thailand	THA	2023	15.69	0.64

Table S11: Reporter-level tariff context for exposed goods, using the latest available WITS/TRAINS year in 2023-2021.



(a) Integration-and-data dependency share.

(b) Core-country spotlight panel.

Figure S9: Supplementary diagnostics for implementation dependency and selected-country structure. These panels help interpret whether cross-country differences reflect only exposure levels or also different implementation bottlenecks.

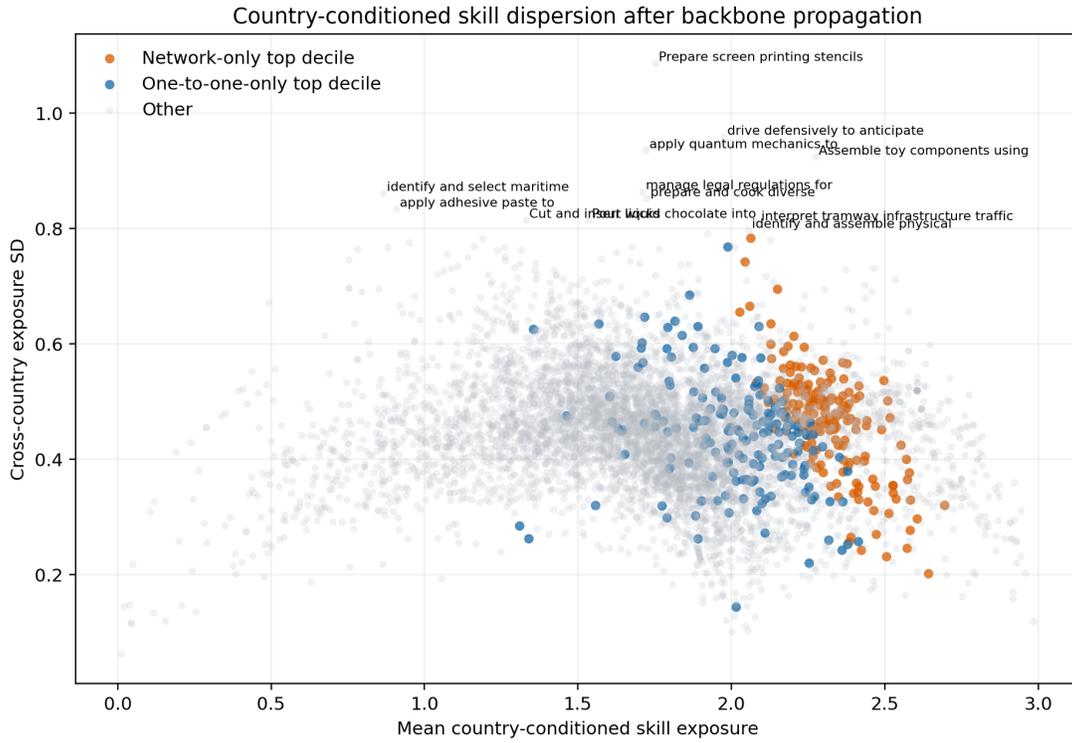


Figure S11: Country-conditioned skill dispersion after backbone propagation. The skill layer shows that country sensitivity does not vanish once task-level heterogeneity is projected through the preserved backbone; some skills combine high average exposure with unusually wide cross-country spread, and some of those same skills are also sensitive to one-to-one compression.

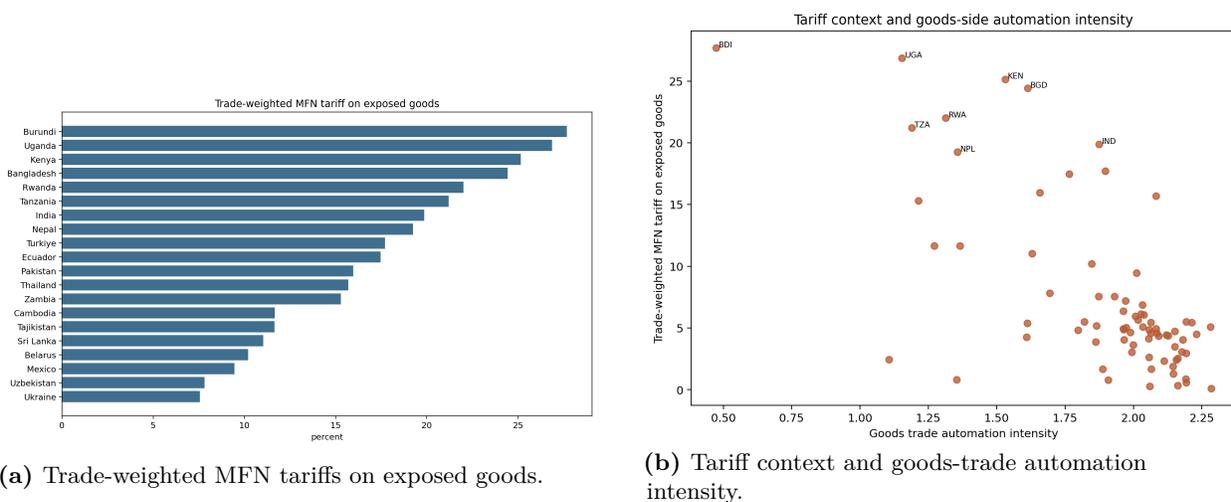


Figure S12: Tariff context figures from the WITS/TRAINS Supplementary Materials layer.

be treated as a substantive empirical analysis rather than an illustrative add-on. The goal is not exhaustiveness for its own sake; it is to show that the trade-facing analysis rests on a sizeable and interpretable linkage structure.

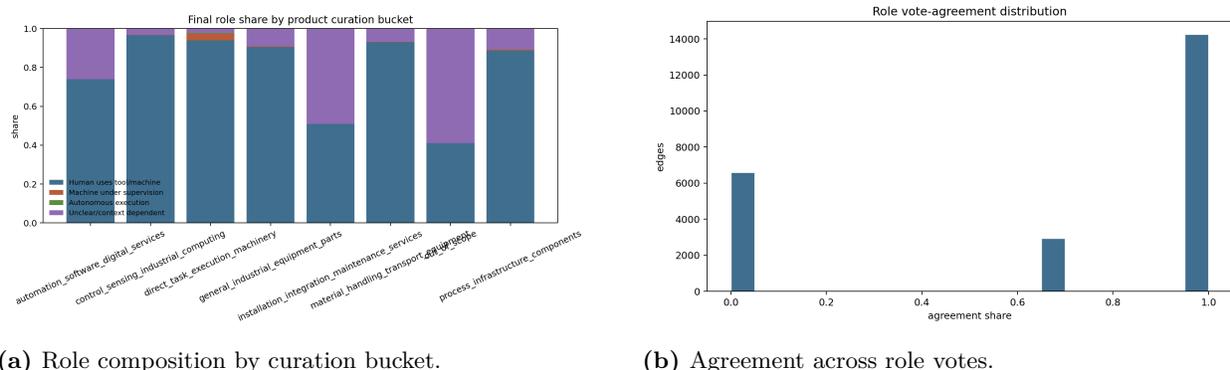


Figure S13: Product-role diagnostics for the subset with explicit role labels. The panels summarize both composition and agreement, which matters because the product branch is meant to refine interpretation rather than simply enlarge coverage.

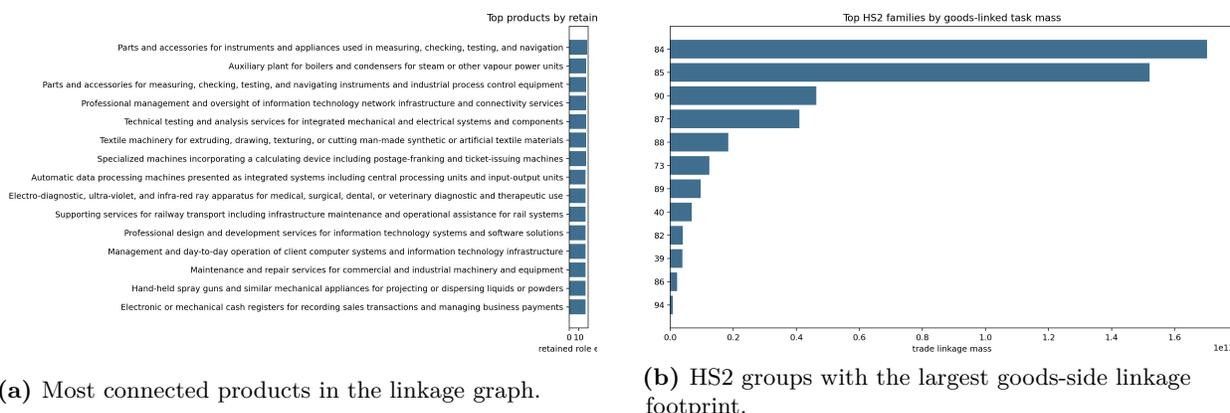
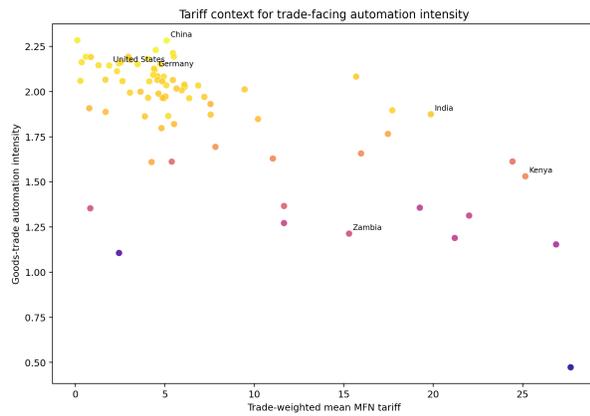


Figure S14: Additional product and trade-linkage diagnostics. These panels show where the linkage graph is most concentrated and which parts of the product space carry the largest task-facing footprint.

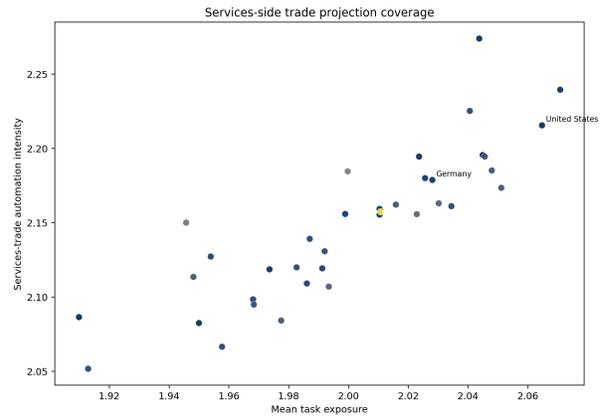
These Supplementary Materials results also help define the empirical boundary of this paper. BACI and WITS/TRAINS are included here. BaTIS, OECD TiVA/ICIO bulk files, and BTDiXe would broaden the services and value-chain side in future work, but they are not required for the present paper.

References

- Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.
- Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.
- Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.



(a) Tariff context and trade-facing automation intensity.



(b) Services-side trade projection coverage.

Figure S15: Additional trade-country mechanism diagnostics. These figures extend the main mechanism argument into tariff-linked and services-side domains while keeping the interpretation descriptive rather than causal.

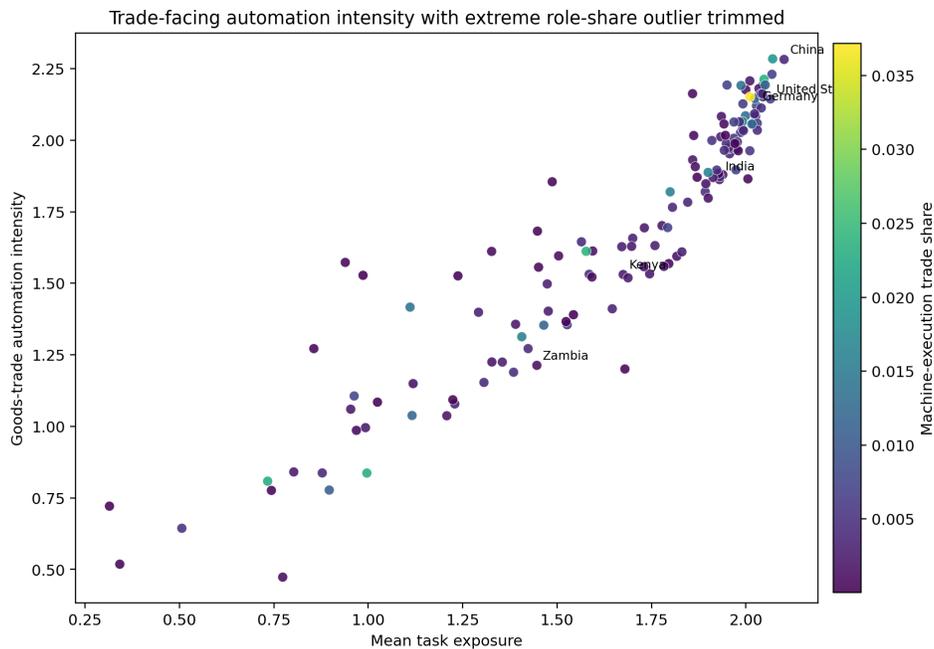


Figure S16: Trimmed version of the country exposure and trade-facing automation-intensity scatter. To show the central cross-country pattern more clearly, this display trims countries above the 99.5th percentile of goods-side machine-execution trade share for plotting. The full-sample figure remains in the main text.

- Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. arXiv preprint.
- Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308.
- Felten, E., Raj, M., and Seamans, R. (2021). Occupational heterogeneity in exposure to artificial intelligence. Working paper; SSRN 3690610.
- Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280.
- Webb, M. (2020). The impact of artificial intelligence on the labor market. *SSRN Electronic Journal*.