Global Automation Atlas*

Prashant Garg
Imperial College London

Tommaso Crosta
Bocconi University

Jasmin Baier
University of Oxford

May 2026

Abstract

Automation affects the labour content of work differently across different contexts. Yet, most existing exposure measures assign fixed scores to tasks or occupations, limiting comparisons of automation exposure across countries. We develop a task-based and country-specific approach to classify automation exposure across the world to disentangle labor-substituting from labor-augmenting automation, the relevant technology channel, and the material role of AI. Our measure spans 124 countries, generating an atlas of 2.33 million task-country labels for economies covering 99% of world population and GDP. We present five descriptive results. First, exposure is highly uneven, ranging from 3.3% of tasks in South Sudan to 61.6% in China, and rises strongly with income, although substantial variation remains within income groups. Second, across countries, exposed tasks are skewed towards substitution rather than augmentation, but low-income countries are disproportionately exposed to substitution, whereas middle-income countries are more heterogeneous. Third, less technologically advanced forms of automation account for more than half of exposed tasks in low-income countries but about one quarter in high-income countries; while other more complex channels generally rise with income levels. Fourth, AI tends to be less prevalent in simpler channels of automation, but also more prevalent in labour-substituting margins in lower income settings and to augment labour in higher income settings. Fifth, we find that females seem to be disproportionately more exposed to labour-substituting automation than males. Our methodology provides a basis for comparing automation exposure across development stages, linking it with cross-country data and allowing us to treat exposure levels, labour margins, technological channels and AI involvement as separate dimensions.

1 Introduction

Automation has become a central object in the study of technological change, work organisation, and labour-market adjustments. Early work on industrial automation has shown that, historically, automation is not a deterministic technological process, but a contested reallocation of tasks, discretion, and control within the production process (Noble, 1984). The central question is therefore not only whether a technology can perform a task, but which functions it can take over, which human roles it leaves in place, and how firms organise that division between workers and machines. This insight runs through modern task-based theory, which models automation as labour displacement from existing tasks, reinstatement through new tasks, and reallocation across newly created ones (Autor et al., 2003; Acemoglu and Restrepo, 2018, 2019; Restrepo, 2024).

*The accompanying online atlas is available at automationatlas.org. It provides country-, occupation-, industry-, and task-level exposure measures, with documentation and downloadable data.

This historical and theoretical work leads to the important need to measure automation, both in its present and future adoption, making "exposure" a central concept of interest. Thus, a growing empirical literature traced automation through specific technological innovations: from computerization and software, to robots, machine learning (ML), and generative AI (LLMs) (Michaels et al., 2014; Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020; Brynjolfsson and Mitchell, 2017; Webb, 2020; Eloundou et al., 2024). Existing measures have advanced the empirical study of automation substantially, each designed around a different measurement unit: occupation-, sector-, country-, or at the task-level (Tomei and Klein Teeselink, 2026). We argue that the task is indeed the most natural unit of measurement, given the insights and results of the theoretical literature. Existing measures have focused on specific *channels* or *features* of technological change, such as computerisation risk (Frey and Osborne, 2017), robot exposure (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020; Agnolin et al., 2025), AI capability alignment (Brynjolfsson et al., 2018; Felten et al., 2021), and ML and LLM exposure (Webb, 2020; Eloundou et al., 2024). Cross-country applications often account for variation through employment weights, sectoral composition, or readiness indices, while leaving the underlying task exposure level fixed (Pizzinelli et al., 2023; Gmyrek et al., 2023; Cazzaniga et al., 2024a). These methodological choices make it harder to create generalizable measures of automation exposure at the task- and country-level, that are moreover able to identify which *functions* of work are targeted by the automation process. A further challenge is that AI is often treated as the channel of automation itself, rather than one of its technological features, operating through several channels and with different margins of labour substitution or augmentation. Measuring automation more broadly requires separating its technological features from the function it performs: the same AI system can operate through workflow execution, planning, information transformation, inference, or physical execution, and each route can substitute labour, augment labour, or reorganise work around a shared human–technology process (Parasuraman et al., 2000; Bertsimas and Kallus, 2020; Raisch and Krakowski, 2021; Amershi et al., 2019).

We thus develop a task-based and country-specific measurement architecture that records all these dimensions jointly. Using the O*NET task dictionary, we classify 18,797 standardised tasks across 124 country settings, covering economies that account for about 99% of world population and GDP (O*NET Resource Center, 2024b,a). For each task-country pair, we record the level of economic exposure, the dominant technology channel, the labour margin, whether AI is materially involved, and, where relevant, the dominant AI function. We are then able to aggregate exposure measures at the country, occupation, and industry level.

Because the labels are produced with Large Language Models, we validate the model in four complementary ways. Supplementary Note Supplementary Note A.4 evaluates *construct validity* by comparing parts of our measurement architecture to the closest existing measures in the literature: LLM exposure in Eloundou et al. (2024), AI exposure in Felten et al. (2021), robot exposure in Webb (2020), country-level AI preparedness in Cazzaniga et al. (2024a), and firm-reported AI adoption from Eurostat. We then test *convergence validity* across an independent model family, *reasoning consistency* across rationales and prompt paraphrases, and *face validity* through direct inspection of label distributions and anchor occupations. Taken together, these checks suggest that the measure recovers established aggregate patterns while adding the more comprehensive and granular decomposition by country, channel, labour margin, and AI materiality.

We report five descriptive results, following the order of the empirical section. First, country-conditioned exposure is highly uneven. The economically exposed share of tasks ranges from 3.3% in South Sudan to 61.6% in China. Exposure rises strongly with per capita income, but income tiers do not exhaust the variation: substantial heterogeneity remains within income groups, especially among middle-income countries. Second, distinguishing labour-substituting from labour-augmenting automation is crucial. Substitution-only exposure is larger than augmentation-only exposure in every country in our sample; low-income countries are especially skewed toward substitution among exposed tasks, while middle-income countries display wider heterogeneity in the substitution-versus-augmentation frontier.

Third, the channels of automation exposure change with income levels, together with the salience of AI across the channels. Less technologically advanced forms of automation account for more than half of exposed tasks in low- and lower-middle-income countries but roughly one quarter in high-income countries, while more complex channels generally become more relevant as per capita income rises. This is intuitive, but underscores and quantifies how solely focusing on AI-related exposure in automation disregards economically meaningful dynamics in the global labour market. Moreover, the share of exposed tasks for which AI is materially involved ranges from roughly 35% to more than 70% across income brackets. It is less prevalent in simpler channels, and its relationship with labour margins changes across countries: in lower-income settings it is more tied to substituting exposure, while in higher-income settings it is more closely associated with augmenting and shared human–technology workflows.

Fourth, aggregating the task-country labels to occupations and industries recovers important patterns at standard labour-market levels. Some occupations and industries are already exposed at lower income levels, especially clerical, transactional, and routine information-processing work; others rise steeply with development, including business administration, ICT, plant operation, manufacturing, and information-intensive services. Complementing the measure with ILOSTAT data shows that females are more exposed to labour-substituting automation both across occupations and industries, while males are more exposed to labour-augmenting automation across industries.

Fifth, we link the task-country measures to country-level data on income, connectivity, schooling, institutions, and capital intensity. GDP per capita, internet use, years of schooling, and regulatory quality are the most informative variables for predicting the exposed-task share. Within exposed tasks, the relevant covariates differ by labour margin: substitution-only composition is lower in countries with higher per capita income, regulatory quality, and capital intensity, while augmentation-only composition is most closely related to capital intensity. These patterns motivate further research on the country-level determinants of automation intensity and margin composition.

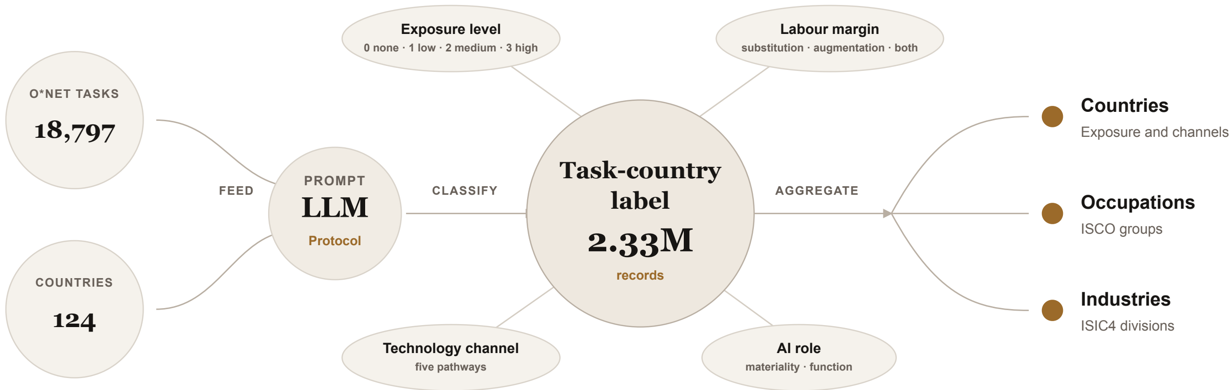
The rest of the paper proceeds as follows. Section 2 defines the data and measurement pipeline: task universe and labels, country conditioning, aggregation methods, country predictors and labour-force inputs, and validation design. Section 3 reports the five descriptive results: trends in global exposure, labour margin frontier, channels of automation and the role of AI, occupation and industry aggregates, and country-level correlates. Lastly, Section 4 discusses what the measure implies for studying the drivers of automation and closely related literatures. The supplementary notes provide construction details, prompt and diagnostic details, construct validity, internal validity, country-conditioning checks, and results companions.

2 Data and Measurement

We measure automation exposure at the level of work tasks. The task dictionary is based on O*NET, a detailed database of occupational work activities created and maintained by the US Labour Department. We take each standardised O*NET task and evaluate it separately in each country context using a structured language-model classification protocol. For each task-country pair, the protocol records exposure level, labour margin, dominant technology channel, AI materiality, and AI function. The resulting dataset contains 2.33 million task-country labels, which we aggregate to country, occupation, and industry measures (Figure 1).

This section outlines the O*NET task universe, the classification protocol, and the country-

Figure 1: Constructing task-country exposure measures.



Notes: The figure summarises the task-country labelling pipeline. We combine 18,797 O*NET task statements with 124 country contexts and classify each task-country pair under a fixed labelling protocol. Each record stores an exposure level, labour margin, dominant technology channel, AI materiality, and AI function. Exposure is coded on a 0–3 scale; levels 2 and 3 define the economically exposed share. Labour margin distinguishes substitution-only, augmentation-only, and balanced-both cases, where both routes are materially plausible. The dominant channel is one of five implementation pathways: physical execution, rule-based workflow, planning/control, information transformation, or inference/scoring. AI materiality records whether AI/ML is central to the automation route; AI function is recorded when AI is material. The resulting 2.33 million retained records are aggregated to country, occupation, and industry measures.

conditioning design. It then describes the benchmark variants, aggregation links to occupations and industries, additional cross-country variables used in the analyses, and measurement robustness checks. Supplementary Table A.1 lists the main data objects.

2.1 Task source and automation labels

We use O*NET version 29.1, released in November 2024, and retain 18,797 unique standardised task statements and their occupation links (O*NET Resource Center, 2024b,a). These tasks form the labelling universe for the language-model classification. We use tasks as the unit of measurement because occupation titles can bundle different work activities across workplaces and countries; country, occupation, and industry summaries are then built by aggregating task-level labels. This choice follows the task-based literature on technological change, where automation operates by reallocating task content between workers and technology (Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018).

Task automation labelling uses Gemini 3.1 Flash-Lite under a fixed structured classification protocol. We apply the country-conditioned version to 124 country contexts, spanning economies that account for 99.0% of world population and 99.1% of world GDP. The retained dataset contains 2,330,776 country–task observations.

Each country–task observation is labelled along five dimensions. The *exposure level* records how much of the task can plausibly be automated with currently available technology. The *dominant technology channel* records the mechanism through which automation reaches the task core. The *AI materiality* flag records whether contemporary AI/ML models are central to that mechanism; when they are, the *dominant AI function* records the role they play. Finally, the *labour margin* records whether the dominant route is substitution, augmentation, or a balanced combination of the two. The same protocol is also run under income-group and context-free settings for benchmark comparisons. The exact prompt schemas and model configurations are reported in Supplementary

Note A.2.

We keep these dimensions separate because they answer different empirical questions. Exposure records the *extent* of economically credible automation; channel records the *mechanism* through which automation reaches the task core; labour margin records the *labour-reallocation route*; and AI materiality records whether AI/ML is central to that route. Country context enters as the assessment setting: the same task is evaluated under the production conditions, complements, and constraints of the named country.

This schema also clarifies how the measure relates to earlier exposure instruments. Capability-alignment measures emphasise what AI or machine-learning systems can do (Felten et al., 2021; Brynjolfsson et al., 2018); text-similarity measures connect technologies to occupational or task descriptions (Webb, 2020); and exposure-intensity measures summarise automatable work into a single score (Frey and Osborne, 2017; Eloundou et al., 2024). Our schema keeps these margins visible in the same task-country measurement exercise.

Exposure levels. Exposure is defined as whether currently available technology can perform, transform, or materially reorganise a nontrivial share of the task core at sufficient quality, reliability, and effective cost. This definition follows the task-based view that automation reallocates task content between labour and technology (Autor et al., 2003; Acemoglu and Restrepo, 2018). The four-level scale separates no credible automation margin, assistive contact, partial economic exposure, and extensive economic exposure:

- **Level 0**. *no credible economic automation margin*: current technology cannot create a credible labour-reallocation margin for the task core.
- **Level 1**. *assistive-only contact*: technology may help, but does not materially reduce human labour input on the task core.
- **Level 2**. *meaningful partial economic exposure*: a nontrivial share of the task core can be reassigned away from labour, while humans remain central.
- **Level 3**. *extensive economic exposure*: a large share of the task core can be reassigned away from labour, or the task is mostly automatable in typical settings.

Throughout the paper, the *economically exposed share* means the share of tasks at levels 2 or 3, and the *high-exposure share* means the share at level 3 alone. Level 1 is an assistive-contact category: it records cases where technology may help with the task but does not create an economically meaningful automation margin. The cutoff keeps broad tool assistance separate from task-core automation that creates an economically material labour-reallocation margin (Autor et al., 2003; Acemoglu and Restrepo, 2018). It also reflects the human-factors point that automation can leave people responsible for monitoring, oversight, and exception handling even in heavily automated settings (Bainbridge, 1983; Endsley, 2017).

Dominant technology channel. Each labelled task is assigned one dominant channel from a fixed taxonomy of five technology archetypes, with *none* reserved for cases where no automation mechanism is identified:

- **Physical execution**: robotics, mechanical actuation, and embodied systems that physically perform the task, such as warehouse picking or assembly automation (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020).
- **Rule-based workflow**: scripted software, robotic process automation, and deterministic business rules that execute structured workflows without learning, such as claims routing or payroll

rules (Autor et al., 2003; Syed et al., 2020).

- **Information transformation**: systems that extract, translate, transcribe, summarise, or generate informational content when the transformed content is the economically relevant output, such as transcription, translation, drafting, or form extraction (Eloundou et al., 2024; Brynjolfsson et al., 2025).
- **Planning/control**: optimisation, scheduling, routing, and control systems that choose or update allocations, action sequences, or control settings under objectives and constraints, such as dispatch, routing, or scheduling (Bertsimas and Kallus, 2020; Parasuraman et al., 2000).
- **Inference/scoring**: statistical or learned models that predict, classify, detect, rank, or score from data when the assessed output is economically relevant, such as risk scoring, defect detection, or document classification (Brynjolfsson et al., 2018; Webb, 2020).

The channel is assigned by the mechanism that produces the task's economically relevant output, not by tool names, upstream models, or sector labels. This mechanism-first rule follows automation and human-factors work that classifies automation by function and control stage (Endsley and Kaber, 1999). If an inferential model only informs scheduling, routing, or allocation, the channel is *planning/control*; if extraction or summarisation feeds a deterministic workflow, the channel is *rule-based workflow*; and if learned perception is embedded in a machine whose contribution is physical actuation, the channel is *physical execution*. The taxonomy therefore separates tasks where inference or information transformation is the task-core output from tasks where learned models support a broader workflow, planning, or physical-execution system.

Labour margin. The labour-margin label is assigned separately from the exposure level and only among tasks that clear the economic-exposure threshold. It records the dominant labour-reallocation route within economically exposed work: substitution-only, augmentation-only, balanced-both, or unclear. Substitution-only means technology can reduce human labour input on the task core. Augmentation-only means technology mainly raises worker productivity while humans remain central to the task. Balanced-both is reserved for cases in which both routes are materially plausible, and the route taken primarily depends on worker capability or work organisation and management. The *unclear* category denotes genuine measurement uncertainty.

Each task-country record that clears the exposure threshold is assigned to one margin. Country margin shares use the full task universe as the denominator unless explicitly labelled within-exposed: a substitution-only share is the share of all tasks with exposure level 2 or 3 and a substitution-only margin. Non-exposed (level 0) and assistive-contact (level 1) tasks remain outside the three exposed-margin shares, while within-exposed shares renormalise the same exposed tasks to sum to one.

This distinction follows task-based models in which technologies can either displace labour from task content or raise the productivity of remaining workers (Acemoglu and Restrepo, 2018). It also aligns with management work that treats automation and augmentation as distinct organisational designs (Raisch and Krakowski, 2021). Recent field evidence on generative AI adoption shows why the distinction matters: the same exposed task can look substitution-heavy or augmentation-heavy depending on workflow, deployment choice, and division of labour (Brynjolfsson et al., 2025; Noy and Zhang, 2023).

AI materiality and dominant AI function. AI materiality records whether contemporary learned models are central to the technology that would perform the exposed task core. When AI is material, the schema also records one dominant AI function:

- **State inference**: learned prediction, classification, detection, or recognition used to infer a relevant state of the world from data (Brynjolfsson et al., 2018; Webb, 2020).
- **Content transformation**: learned drafting, summarisation, translation, extraction, or rewriting where transformed informational content is the material AI contribution (Eloundou et al., 2024; Brynjolfsson et al., 2025; Noy and Zhang, 2023).
- **Recommendation and decision support**: learned ranking, prioritisation, proposal, or advice that supports human or organisational choice without itself constituting the final task output (Puranam, 2021; Amershi et al., 2019).
- **Adaptive control**: learned updating of control actions or control-relevant parameters inside embodied or robotic systems (Guo and Pan, 2023).

AI materiality is separate from channel because many exposed tasks are automated through legacy software, optimisation, or robotics, with learned models playing at most a supporting role. This dimension asks whether AI is central to the exposed automation route; the dominant AI function records what the learned model does within that route. This contribution-based treatment follows work that separates the role of AI inside a larger system from the broader process being automated and from the residual human role around it (Amershi et al., 2019; Puranam, 2021; Raisch and Krakowski, 2021).

Country-level dimension. Each task is evaluated in a named country context, meaning the institutional, productive, and deployment environment of that country. The task-country dataset covers 124 countries, spanning economies that account for 99.0% of world population and 99.1% of world GDP. We treat country conditioning as the baseline because the same nominal task can imply different feasible labour-saving margins across economies when complementary capital, skill mix, production organisation, infrastructure, technology diffusion, and distance to frontier differ (Caselli and Coleman, 2001; Comin and Hobijn, 2010; Acemoglu et al., 2006; Caunedo et al., 2023). Countries are also grouped into four World Bank income levels (low, lower-middle, upper-middle, high) using the analytical classification for fiscal year 2025 (World Bank, 2024a); these groups are used for the income-conditioned benchmark and for the income-gradient figures.

The context-free and income-group runs are used as aggregate benchmarks. They show how the same task labels change when the schema is applied with no country context, with broad development context, or with full country context. The context-free benchmark approximates the single global task assessment implicit in much prior exposure work. The income-group benchmark captures the part of country context summarised by broad development level. Comparing both to the task-country dataset separates common task rankings, income-gradient shifts, and remaining country-level variation. Across all three runs, only the context supplied to the classifier changes: the schema, response structure, and aggregation logic are held fixed. Supplementary Figure A.1 reports the benchmark ladder and shows how exposure patterns change as context is added.

2.2 Task linkages to occupations and industries

Once task labels are defined, we aggregate them to occupation- and industry-level measures. Occupation summaries first use the O*NET task-to-occupation mapping, which links each task to U.S. Standard Occupational Classification (SOC) occupations. This preserves the task-to-occupation structure in the source data, while holding the task-to-occupation bundle fixed across countries. We then map SOC occupations to the International Standard Classification of Occupations 2008 (ISCO-08), the occupation taxonomy used in cross-country labour-force data. Industry summaries require a separate task-to-activity linkage because O*NET does not provide an equivalent industry mapping. The country-specific variation in these summaries comes from the task-country labels. Holding the aggregation bridges fixed keeps the occupation and industry comparisons consistent across countries.

Occupation linkage: task → SOC → ISCO. Occupation summaries are reported at the two-digit ISCO-08 level, the International Labour Organization's occupation taxonomy for cross-country labour statistics (ILO, 2012). O*NET links tasks to U.S. SOC occupations, which we map to ISCO-08 using the SOC–ISCO crosswalk reference maintained by the U.S. Bureau of Labor Statistics (BLS, 2024c). The baseline bridge weights SOC occupations by their linked task content; a modal bridge is retained as a sensitivity check. This type of occupation bridge is standard in cross-country AI-exposure work that maps U.S. occupation or task information into ISCO-based labour-force data (Pizzinelli et al., 2023; Cazzaniga et al., 2024b; Gmyrek et al., 2023). Full bridge details are reported in Supplementary Note A.1.

Industry linkage: task → ISIC4. Industry summaries are reported at the two-digit level of the International Standard Industrial Classification of All Economic Activities, Revision 4 (ISIC Rev. 4), the United Nations taxonomy of economic activities (UN Statistics Division, 2008). ISIC classes describe productive activities carried out by establishments. Because O*NET is occupation-native, industry reporting requires a direct task-to-industry link: which O*NET tasks are meaningful components of each ISIC class's activity.

We construct a task-to-ISIC4 graph in which each retained edge links an O*NET task to an ISIC4 class. The graph uses a bottom-up candidate-then-prune design, following the build-prune logic in Fetzer et al. (2024). Embedding similarity proposes candidate task–ISIC4 edges, and a structured LLM voter removes candidates that share language without representing a meaningful activity component. The retained graph contains 12,294 task–ISIC4 edges across 418 retained ISIC4 classes and 88 two-digit divisions. For the two-digit industry figures, country-task exposure is first averaged within retained four-digit classes and then averaged equally to two-digit divisions and income-group cells. The equal-weight design makes the ISIC results a measure of class content. Full construction details are reported in Supplementary Note A.1; the occupation and industry rankings that use the linkage are reported in Supplementary Note B.5.

2.3 Country predictors and labour-force inputs

Country-level variables. The descriptive exercise in Section 3.5 uses country-level variables from the Penn World Table and other public sources to ask which national characteristics are associated with the amount of exposed work and with the substitution-versus-augmentation mix within exposed tasks. The covariates cover five broad domains. First, income level is measured using log GDP per capita, computed from World Bank GDP and population fields (World Bank, 2024b). Second, human-capital capacity is measured using the Penn World Table human-capital index, which combines schooling and returns to education, and average years of schooling for adults aged 15–64 from the Barro–Lee educational-attainment data (Feenstra et al., 2015; Barro and Lee, 2013). Third, capital deepening and investment are measured using log real capital stock per worker from the Penn World Table and gross fixed capital formation as a percentage of GDP from the World Development Indicators (Feenstra et al., 2015; World Bank, 2024b). Fourth, digital connectivity is measured by the share of individuals using the Internet from the World Development Indicators (World Bank, 2024b). Fifth, institutional and market conditions are measured using government effectiveness and regulatory quality percentile ranks from the Worldwide Governance Indicators, and goods-trade openness, measured as CEPII BACI merchandise trade divided by GDP (World Bank, 2024c; Gaulier and Zignago, 2010). For time-varying covariates, we retain the most recent non-missing value per country in the available window. B.4 provides more details on these inputs.

Predictor ranking. We fit random forests to allow nonlinearities and interactions among covariates that are partly overlapping proxies for development, connectivity, institutions, and production structure. The main figure ranks covariates by mean absolute TreeSHAP values, which express each covariate's average predictive contribution in the same units as the outcome (Lundberg and Lee, 2017; Lundberg et al., 2020). TreeSHAP is used because unconditional permutation-importance rankings can be sensitive when predictors are highly correlated (Strobl et al., 2008). The ranking should still be read by predictor clusters rather than as separate marginal effects. Direction markers come from one-dimensional accumulated local effects in the fitted forest (Apley and Zhu, 2020). Supplementary Note B.6 reports wider-coverage specifications, the earlier permutation-importance version, variance-inflation diagnostics, and a linear Shapley R^2 decomposition.

Labour-force inputs. ILOSTAT provides labour-force, occupation, and industry employment data used to reweight the constructed exposure measures (ILO, 2026; ILOSTAT, 2026a,b). These data enter after the task-country labels have been constructed. For employment-weighted occupation summaries, we use the most recent 2018–2024 employment shares by country and ISCO-08 major group. For the sex-specific employment analysis, we use female and male employment shares at the most granular public cross-country level available for both views: two-digit ISCO-08 occupation groups and two-digit ISIC Rev. 4 industry divisions. These shares reweight the occupation and industry exposure summaries, so the female–male comparison holds exposure scores fixed and varies observed employment composition.

2.4 Measurement validity

We assess the task labels through four measurement checks, summarised here and reported in detail in Supplementary Note A.4 and Supplementary Note A.5. The checks assess both external alignment and internal consistency. First, we ask whether aggregates built from the labels align with independent measures along the dimensions they are meant to capture. We then examine cross-model convergence, rationale and prompt consistency, and the distribution of labels across tasks and countries.

Construct validity. Construct validity asks whether aggregates built from the labels line up with external measures, often derived from primary data, that were not used to produce them. These include occupation-level exposure scores from Frey and Osborne (2017), Webb (2020), Felten et al. (2021), and Eloundou et al. (2024); the IMF AI Preparedness Index (Cazzaniga et al., 2024a); and firm-reported AI adoption from the Eurostat ICT Usage in Enterprises AI module. Because these comparators measure different parts of exposure, the test is construct-specific: agreement should be strongest along the matched dimension. We compare AI-focused measures with AI-material share, robotics measures with physical execution, and foundation-model measures with inference and information-transformation channels.

The external comparisons follow that pattern. Eloundou et al. score tasks and occupations by exposure to GPT-style systems; their broad GPT-4 gamma occupation score is therefore closest to our foundation-model-like share, defined as the sum of inference/scoring and information-transformation channels. This matched share correlates more strongly with Eloundou's GPT-4 gamma score than overall exposure does (Pearson 0.78 versus 0.22 for overall exposure). Felten's

AI Occupational Exposure index measures occupational exposure to AI capabilities, so its closest counterpart in our schema is AI-material share (0.61 versus 0.10 for overall exposure). Webb's robotics-patent sub-score links occupation tasks to robotics patents, so its closest counterpart is physical-execution share (0.72 versus 0.05 for overall exposure), the strongest clean channel correspondence in the matrix (Supplementary Figure A.2).

We further extend this exercise to widely used country readiness and firm adoption data. The IMF AI Preparedness Index summarises country readiness for AI through digital infrastructure, human capital and labour-market policies, innovation and integration, and regulation and ethics. Country-level AI-material share correlates strongly with the IMF index, at Pearson 0.90 across 117 countries. Because the IMF index is built from development-related preparedness components, we also ask whether the alignment remains after removing the common log-GDP gradient. The residual association remains positive, with partial Pearson 0.42 (Supplementary Figure A.3). Eurostat's ICT Usage in Enterprises module reports the share of firms using AI technologies by country and industry. Industry-level AI-material share correlates positively with that reported adoption, at Pearson 0.41 across reported country–NACE cells and 0.52 across reported NACE-cell means (Supplementary Figure A.4). Taken together, these correlations suggest that our labelling recovers meaningful patterns observed in the real world.

Convergent validity. Convergent validity asks whether the task ordering is stable across model families, a central concern in recent multi-model replications of LLM-generated occupational exposure scores (Yin et al., 2026). We re-label the full 18,177-task context-free universe with `gpt-5.4-mini` under the same prompt and schema, and compare the two runs task by task. The four-level ordering is relatively stable: 95.0% of tasks agree within one level. The exact exposed-versus-not-exposed cutoff is more demanding, with 75.1% binary agreement and the independent model drawing the cutoff somewhat lower than the main run (Supplementary Figure A.5). This pattern motivates our robustness analyses around exposed cut-off point to ensure conclusions are not driven primarily by the exact numeric exposure level classification.

Reasoning consistency. Reasoning consistency asks whether the label is recoverable from the model's stated rationale, and whether prompt paraphrases change the classification. We use two tests. First, an independent model receives the task description and original rationale, with the original label withheld, and predicts the label under the same schema. Second, the main labelling model classifies the same task sample under three paraphrased prompts. The independent model recovers the binary exposed label for 99.0% of a stratified 1,000-observation sample from the rationales alone (Supplementary Figure A.6); the three paraphrases leave the exposure level within one step of agreement for 99.8% of tasks (Supplementary Figure A.7). Disagreements concentrate at fuzzy boundaries and rarely cross the exposed-versus-not-exposed cutoff used in the main results.

Face validity. Face validity asks whether the labels look plausible before they are compared with external measures. We use four checks. First, the distributional check asks whether the label shares across all 2,330,776 country-task observations have sensible mass across exposure levels, channels, margins, and AI materiality, both overall and by income group. Second, we explore rationale-label relationships for internal consistency. Third, country-conditioning rationales for the same task are compared across countries to see whether they remain task-relevant while reflecting different production environments, skill levels, and adoption constraints. Fourth, we observe whether occupations with strong prior rankings in the literature, such as clerical, professional, and physical-production roles, appear in the expected order.

The appendix reports the evidence from these checks. The distributional figures show how label shares vary overall and by income group, rather than concentrating in a single category (Supplementary Figures A.9 and A.10). The rationale-consistency screen flags potential rationale–label mismatches in 0.11% of observations (Supplementary Table A.8). The country-conditioning check compares rationales for the same task across country contexts, while the anchor-occupation check asks whether occupations with strong prior rankings appear in the expected order. For example, clerical and administrative occupations sit high in exposure, while occupations centred on in-person care and physical services sit lower. Supplementary Figure A.8 and Supplementary Tables A.9 and A.7 report the supporting examples and rankings.

3 Results

We organise the results around five questions. First, how uneven is automation exposure across countries, and how closely does it track income per capita? Second, once tasks are exposed, do they mainly substitute labour, augment labour, or combine both margins? Third, which technology channels carry exposed work, and when is AI materially involved? Fourth, how do the task-country labels aggregate to the occupation and industry units used in labour-market and production data? Fifth, which country characteristics correlate with exposure levels and with the composition of exposed work?

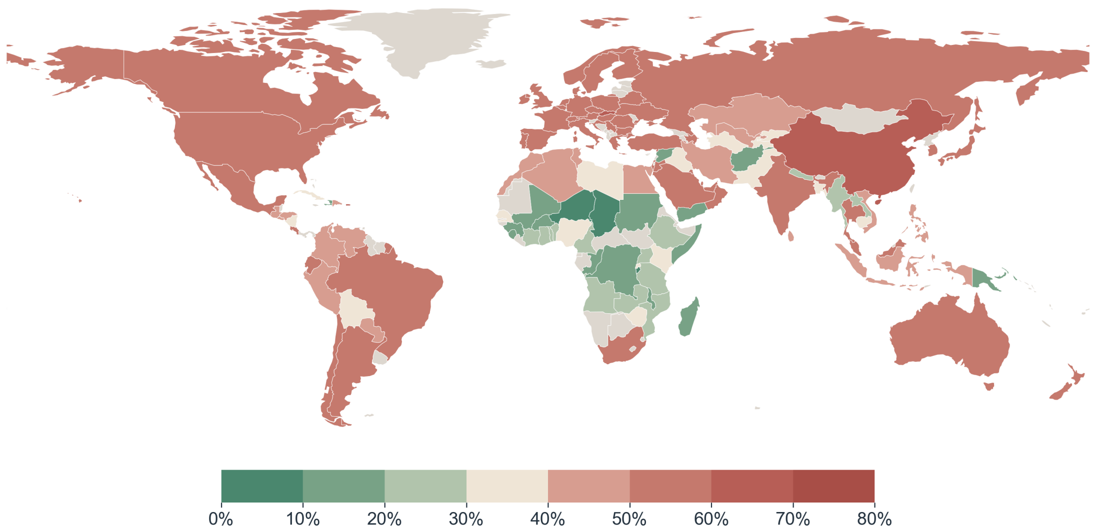
3.1 Global trends in automation exposure

Figure 2 shows that automation exposure is globally uneven and closely associated with, but not fully explained by, income per capita. Panel (a) maps the economically exposed share of tasks: the share rated 2 or 3 on the 0–3 exposure scale, where currently available technology creates a meaningful partial or extensive labour-reallocation margin in the named country context. Panel (b) relates that exposed share to log GDP per capita.

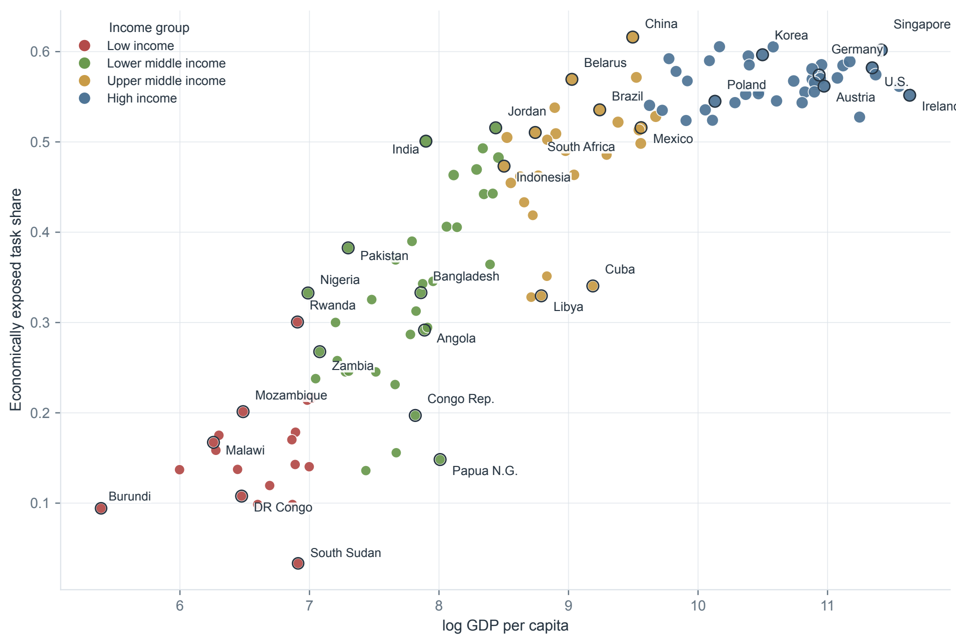
Exposure varies widely across countries, from 3.3% of tasks in South Sudan to 61.6% in China. The exposed task share rises strongly with country income: richer countries have more economically exposed tasks on average. Even within the same income tier, meaningful heterogeneity remains, especially within middle-income countries, consistent with different pathways of structural transformation (Supplementary Figure B.1).

Figure 2: Cross-country patterns in automation exposure.

(a) Economically exposed task share.



(b) Exposed task share and GDP per capita.



Notes: Panel (a) maps the share of tasks rated 2 or 3 on the 0–3 exposure scale, corresponding to meaningful partial or extensive economic exposure. Panel (b) plots the same exposed-task share against log GDP per capita; point size scales with the share of tasks at exposure level 3, and colour denotes World Bank income group.

3.2 Heterogeneity in automation margins

The same exposed task need not imply the same labour-market margin. In some cases, technology can reduce the need for human labour on the task core; in others, it can raise worker productivity while leaving humans central; and in some cases both routes are plausible. We label these three cases substitution-only, augmentation-only, and balanced-both.

Across the full task universe, substitution-only exposure exceeds augmentation-only exposure in every country, but the size of the gap varies substantially and dispersion widens within middle-income countries (Figure 3a). Following the same tasks across income groups, many tasks move from non-exposure into balanced-both and substitution-only margins as income increases, while augmentation-only exposure remains comparatively small (Figure 3b). [1]

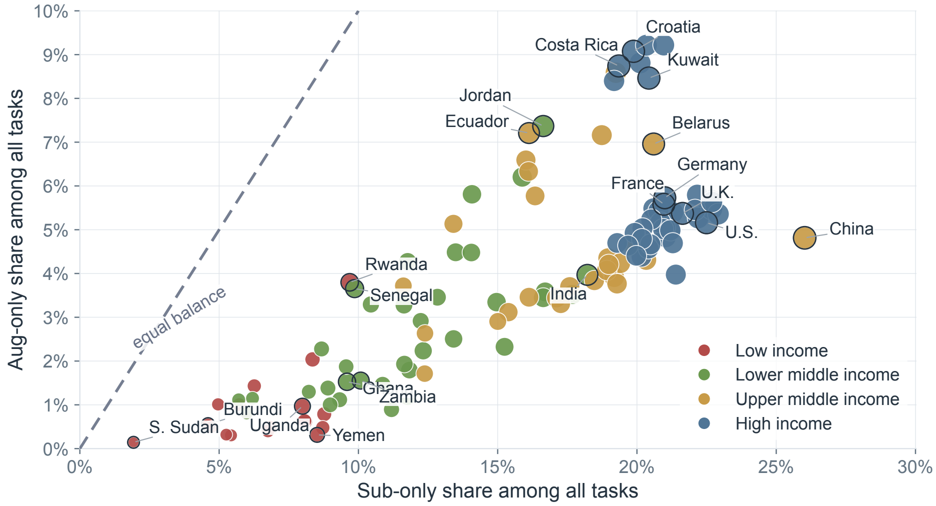
Within exposed tasks, countries differ in how far exposure leans toward substitution or augmentation (Figure 3c). Three patterns stand out: (i) most low-income countries sit in a substitution-heavy part of the automation frontier, with few purely augmentation-oriented paths; (ii) most high-income countries lie closer to the middle of the distribution, although several smaller high-income economies sit near the top of the augmentation distribution; and (iii) lower- and upper-middle-income countries show the widest dispersion, consistent with differences in production structure and structural-transformation paths.

Supplementary Figure B.3 reports margin polarisation, defined as the share of exposed tasks outside the balanced-both category, and its within-income-group dispersion. Polarisation is highest in low-income countries, where substituting-margin exposure is especially strong; declines for lower-middle-income countries; and then plateaus among upper-middle- and high-income countries. The balanced-both margin remains substantial throughout, covering roughly 45% to 55% of exposed tasks. This is consistent with task-based accounts in which automation can displace labour from parts of a task while preserving or raising the productivity of remaining human work (Acemoglu and Restrepo, 2018; Raisch and Krakowski, 2021). It also suggests that for most tasks completed by human labour, work organization and individual decisions may matter more for whether a technology is used in a substituting or augmenting fashion, than the technology itself.

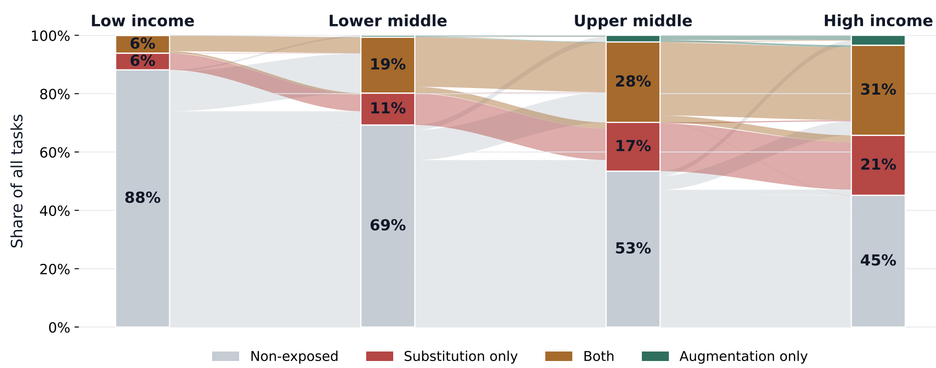
[1] The adjacent-step transition probabilities behind the alluvial ribbons are reported in Supplementary Figure B.2.

Figure 3: Labour-margin composition across countries and income groups.

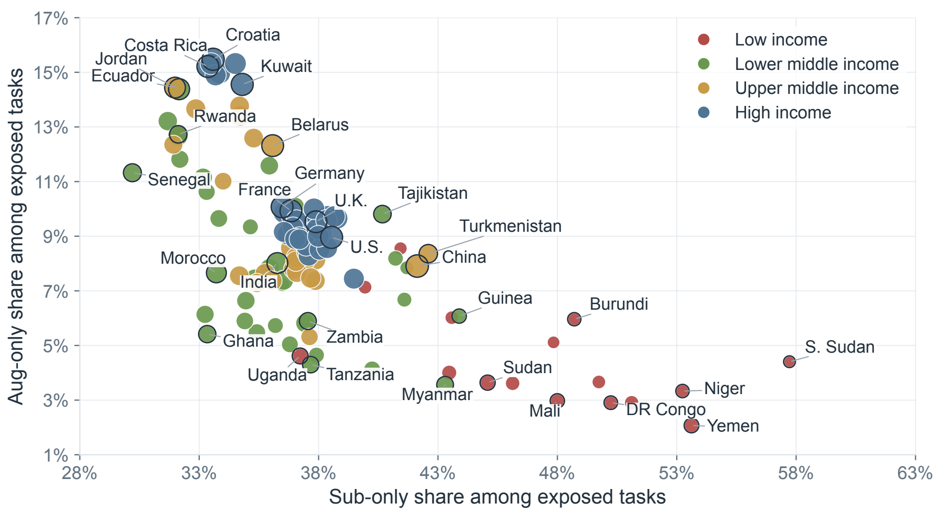
(a) Shares of all tasks.



(b) Same tasks across income groups.



(c) Shares within exposed tasks.

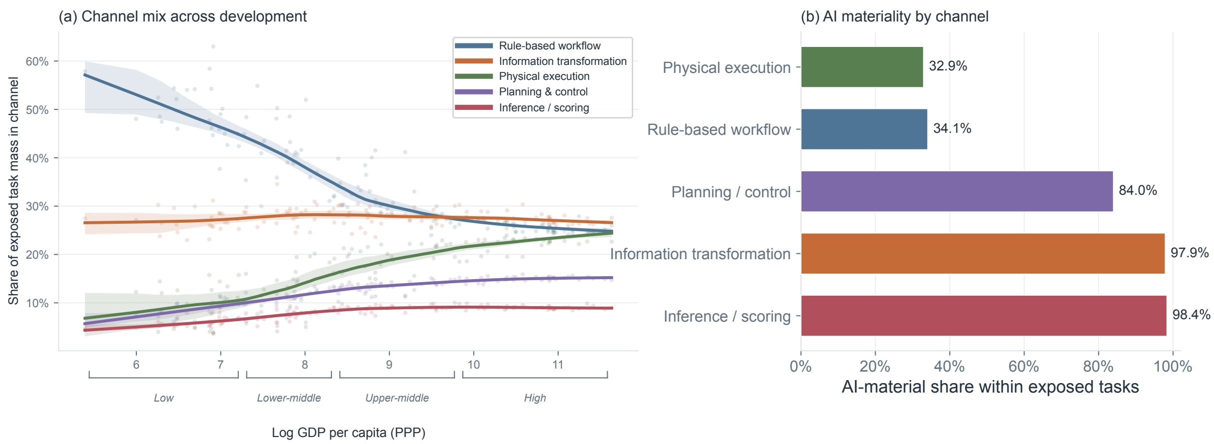


Notes: Panel (a) reports substitution-only and augmentation-only shares as shares of all tasks. Panel (b) follows the same 18,797 tasks across income groups, assigning each task to its modal pathway within each group. Panel (c) reports substitution-only and augmentation-only shares within exposed tasks; the remaining exposed share is balanced-both. Scatter points are countries; colour denotes World Bank income group, bubble size denotes exposed share, and the dashed line marks equal substitution-only and augmentation-only shares.

3.3 Channels of automation and the role of AI

Automation exposure can be carried by different technology channels, and AI can have a material role within several of them.[2]

Figure 4: Automation channels and AI materiality.



Notes: Panel (a) shows the country-level share of exposed tasks in each dominant technology channel against log GDP per capita; fitted curves are locally smoothed trends (LOESS) with 95% bootstrap intervals, and brackets mark World Bank income tiers. Panel (b) reports, within each dominant channel, the share of exposed tasks for which contemporary learned models are material to the automation route. Channel categories are mutually exclusive; AI materiality is measured separately and can appear within any channel.

Figure 4a shows a clear trend in the channels carrying exposed work. Rule-based workflow automation accounts for more than 50% of exposed tasks in low-income countries, about 40% in lower-middle-income countries, and about 25% in high-income countries. It remains the largest channel outside high-income countries, even though Figure 4b shows that it has relatively low AI materiality. Physical execution moves in the opposite direction, rising from roughly 5% of exposed tasks in low-income countries to about 25% in high-income countries, also with low AI materiality. Information transformation is highly AI-material but accounts for a nearly constant share of exposed tasks across income groups. Planning/control and inference/scoring rise more gradually with income and are also more AI-material.

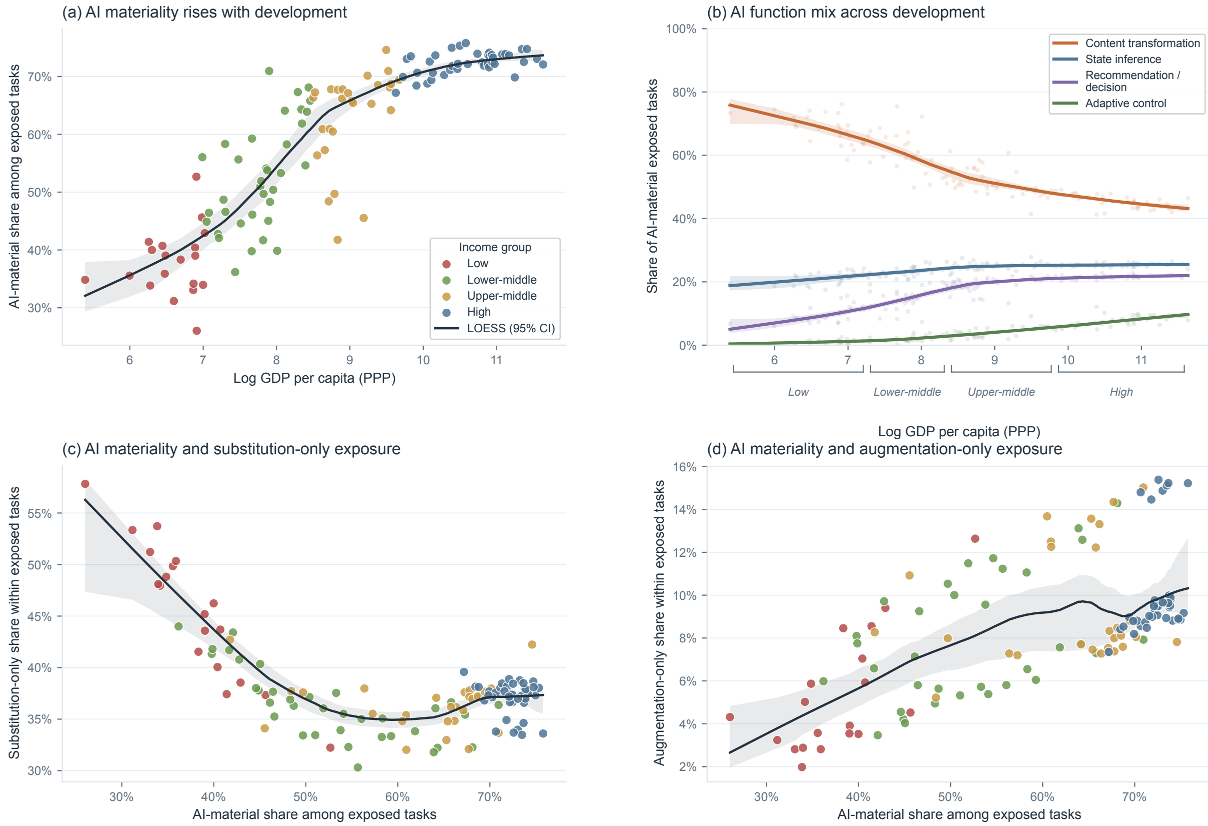
This pattern is why we treat AI as a feature of automation rather than as the channel itself. A large share of exposed work still travels through rule-based workflow systems and physical execution, while richer economies carry a more diversified mix of cognitive and machine-facing automation. The rotation is consistent with technology-diffusion and task-content evidence across development stages (Comin and Hobijn, 2010; Lewandowski et al., 2025).[3] Supplementary Figure B.5 shows the same rotation on an all-task denominator: rule-based workflow exposure rises in levels but plateaus by the upper-middle-income range, while most additional exposed task mass at higher development levels comes from physical execution, planning/control, information transformation, and inference/scoring.

The AI-material share of exposed tasks rises with income, from roughly 35% at the low-income end to more than 70% at the high-income end (Figure 5a). The function mix also changes with

[2]The channel and AI-material breakdowns are composition measures defined after the economic exposure threshold has been crossed.

[3]The within-channel composition, meaning how exposed mass inside each channel splits between substitution-only, balanced-both, and augmentation-only, is broadly similar across the four income groups. Supplementary Figure B.6 reports this breakdown.

Figure 5: AI materiality, AI function mix, and labour-margin composition.



Notes: Panel (a) plots the AI-material share among exposed tasks against log GDP per capita. Panel (b) decomposes AI-material exposed tasks by dominant AI function. Panels (c) and (d) relate the AI-material share among exposed tasks to the substitution-only and augmentation-only shares among exposed tasks. Points are countries; colours denote World Bank income group; fitted curves are locally smoothed trends (LOESS) with 95% bootstrap intervals.

development. Content transformation remains the largest AI function but declines in relative share, while recommendation/decision and adaptive-control functions become more common (Figure 5b). The same pattern appears when countries are grouped by income: lower-income countries are more concentrated in content transformation, while higher-income countries shift toward recommendation/decision and adaptive control (Supplementary Figure B.7).

AI-material exposure also differs by labour margin. Countries with higher AI-material shares among exposed tasks tend to have lower substitution-only shares and higher augmentation-only shares (Figure 5c,d). This pattern follows from the trends in channel rotation above: rule-based workflow and physical execution are less AI-material and more substitution-heavy, while the more AI-material channels are more often tied to augmentation or balanced-both exposure. AI-material exposed work is therefore more likely to sit in shared human–technology workflows than in purely substitutive routes.

3.4 Occupation- and industry-level exposure

The preceding results compare countries using the same task universe. Many labour-market and production datasets, however, are organised by occupation or industry, which bundle tasks into the categories used in employment, wage, production, and trade data. Aggregating the task-country labels to those units shows whether the trend in per capita income survives this bundling, weakens after aggregation, or is concentrated in particular occupational and industrial categories (Autor et al., 2003; Autor and Handel, 2013; Arntz et al., 2016). It also matters for structural-transformation questions, where development is studied through reallocation across occupations, sectors, and tasks (McMillan and Rodrik, 2011; Rodrik, 2016; Caunedo et al., 2023; Lewandowski et al., 2025). We therefore report exposure by ISCO-08 occupation and ISIC Rev. 4 industry, keeping the income-group comparison that structures the country results.[4]

Occupation-level exposure rises across income groups, but unevenly across occupation families (Supplementary Figure B.8). Clerical support occupations are already highly exposed in low-income countries, consistent with the routine-task literature on codifiable clerical and transactional work (Autor et al., 2003; Goos et al., 2014). Business-administration and ICT professional occupations rise sharply with income, aligning with AI-exposure measures that place information-processing and professional-service work high in the exposure distribution (Felten et al., 2021; Eloundou et al., 2024; Gmyrek et al., 2023). Plant-operator exposure also rises with income, consistent with robotics evidence in which physical automation becomes more relevant in richer and more capital-intensive production systems (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020). The margin split sharpens this pattern: substitution-only exposure is highest in transactional and clerical occupations, while augmentation-only exposure is highest in professional, care, and analytical roles (Supplementary Table B.2).

Industry-level exposure shows a similar income pattern (Supplementary Figure B.9). In manufacturing, exposure rises markedly with income in divisions such as food, textiles, wood, paper, and chemicals, consistent with evidence that robotics and ICT reshape industry-level labour demand and task composition (Graetz and Michaels, 2018; Michaels et al., 2014). In services, exposure rises in wholesale trade, publishing, information, legal and accounting activities, advertising, employment services, and office support. These service patterns mirror the high exposure of information-intensive administrative and professional work in occupation-based AI measures (Felten et al., 2021; Eloundou et al., 2024; McElheran et al., 2024). Supplementary Table B.3 reports the top substitution-only and augmentation-only divisions by income group.[5]

The occupation and industry linkages also allow the task-country labels to be reweighted by sex-specific employment shares. This asks whether gender exposure gaps are stable across two standard ways of observing labour markets: occupations, which group workers by the type of work they do, and industries, which group employment by production activity. Recent gender analyses of AI exposure find that women are often more represented in AI-exposed occupations, including six-country evidence from Cazzaniga et al. (2025) and European evidence linking AI exposure to rising female employment shares within such occupations (Albanesi et al., 2025). Our ILOSTAT linkage shows that the answer changes with the weighting frame. Under occupation weights, substitution-only exposure is more female-weighted in the median country, while augmentation-only exposure is close to balanced. Under industry weights, substitution-only gaps are small, while augmentation-only exposure is more male-weighted (Supplementary Figure B.13). A cell-level fixed-effect check separates this aggregate gap from within-country sorting across occupation and industry cells,

[4] Occupation summaries link O*NET tasks to SOC occupations and then to ISCO-08 using the SOC–ISCO bridge described in Section 2.2. Industry summaries use retained task–ISIC Rev. 4 class links. Values are first computed within countries and then averaged within World Bank income groups. Country-level occupation and industry profiles are available in the accompanying online atlas: `https://automationatlas.org/`.

[5] Supplementary Figure B.9 is a detailed class-content measure. Equal weighting preserves the class structure because ISIC does not publish credible class-size weights at this granularity. Within each country, task exposure is first computed at the retained four-digit class level, then averaged equally to two-digit divisions, and finally averaged equally across countries within each income group. The pathway decomposition that carries the substitution, balanced-both, and augmentation partition into occupation- and industry-level measures is reported in Supplementary Figure B.4; top occupation and industry rankings by labour margin are reported in Supplementary Note B.5.

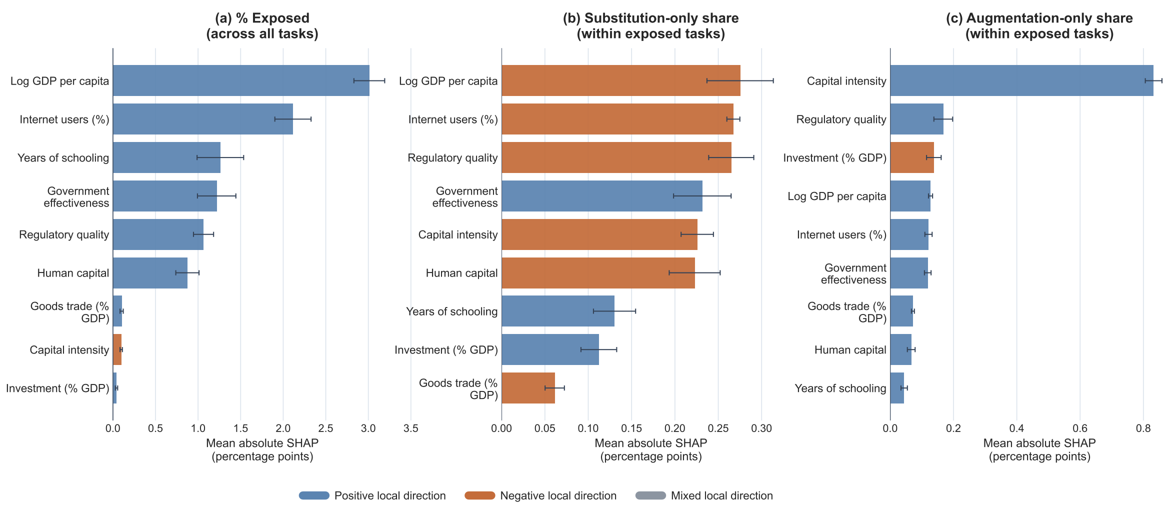
absorbing both country-wide gender-employment differences and average differences across cells. It points in the same direction for substitution: more substitution-intensive cells employ relatively more men, while the corresponding augmentation patterns are weaker and less precisely estimated (Supplementary Table B.7).[6]

3.5 Country-level correlates of automation exposure

The preceding results show that per capita income is strongly correlated with our exposure measure, but it may nonetheless proxy for several other country-level variables: digital connectivity, education levels, institutions, capital intensity, and trade. We therefore ask which country covariates are most correlated with (i) the exposed-task share and (ii) the substitution-versus-augmentation mix within exposed tasks.

Figure 6 ranks these covariates using random forests and mean absolute TreeSHAP values. The bars measure average predictive contribution in percentage points of the outcome share; colours show the accumulated-local-effect direction. Because the covariates are highly correlated, the ranking is descriptive evidence, and should be interpreted as motivating evidence for future research, rather than causal claims.[7]

Figure 6: Country-level correlates of exposure and labour-margin composition.



Notes: Bars report mean absolute TreeSHAP values from random forests estimated on the 68-country complete-case sample. Values are multiplied by 100, so magnitudes are in percentage points of the predicted outcome share. Panel (a) uses exposed share among all tasks. Panels (b) and (c) use substitution-only and augmentation-only shares within exposed tasks and share a common x-axis. Colours indicate the sign of the one-dimensional accumulated local effect. The figure ranks country covariates by predictive importance; it does not estimate separate marginal coefficients.

Panel (a) shows that, for the level of general exposure to automation, GDP per capita and internet use have the largest mean absolute SHAP values, followed by years of schooling, govern-

[6]Supplementary Figure B.12 compares equal-weighted and ILOSTAT employment-weighted occupation exposure across income groups. Supplementary Table B.6 reports the corresponding country-level shifts from equal weighting to employment weighting.

[7]The fitted forests allow nonlinearities and interactions among the country covariates (Breiman, 2001). TreeSHAP values are averaged in absolute value across countries and across five random-forest seeds (Lundberg and Lee, 2017; Lundberg et al., 2020). Direction markers are read from the one-dimensional accumulated local effect of each covariate within the fitted forest (Apley and Zhu, 2020). Supplementary Note B.6 reports variance-inflation diagnostics, wider-coverage random-forest specifications, the earlier permutation-importance version, and a linear Shapley R^2 companion based on dominance-analysis logic (Budescu, 1993).

ment effectiveness, regulatory quality, and human capital. These variables capture the capability environment through which technologies diffuse: human capital affects absorptive capacity, digital infrastructure affects the intensity of use, and managerial, organisational, and institutional conditions shape whether firms can implement technologies at scale (Nelson and Phelps, 1966; Comin and Mestieri, 2018; Andrews et al., 2020; Bloom and Van Reenen, 2007; Cirera and Maloney, 2017).

Panel (b) shows that, within exposed tasks, substitution-only automation is negatively associated with per capita income and regulatory quality, but positively with government effectiveness. Panel (c) shows that augmentation-only automation strongly correlates with capital intensity, followed by a much smaller contribution from regulatory quality. These patterns suggest that the margin of automation depends on the complementary assets surrounding the technology: where capital, organisational capacity, and human–technology coordination are stronger, automation is more likely to be embedded in workflows that raise the productivity of remaining workers rather than simply replace labour input (Bresnahan et al., 2002; Raisch and Krakowski, 2021; Vaccaro et al., 2024; Babina et al., 2024). [8]

These results show that exposure to automation reflects a broader capability environment in which income, digital connectivity, human capital, state and organisational capability jointly determine which automation routes are economically feasible and desirable by market actors, so that countries at similar income levels can be at different points of the automation frontier because they differ in the complements that make particular technologies usable at scale. (Lewandowski et al., 2025; Comin and Hobijn, 2010; Acemoglu et al., 2006; Cirera et al., 2024).

4 Discussion

This paper develops a task-based and country-specific atlas of automation exposure. It extends the task-based automation literature by measuring not only whether work is exposed, but which function is automated, through which channel, with which labour margin, and under which country context (Autor et al., 2003; Acemoglu and Restrepo, 2018; Restrepo, 2024). Across 18,797 O*NET tasks and 124 country settings, we show that exposure is highly uneven across countries, rises with per capita income levels, yet it remains heterogeneous within income groups, and differs sharply by substitution versus augmentation margins, technology channels, and presence of AI. The unexplained heterogeneity, both within levels and margins, points to the need for future research to investigate the economic, social and political drivers of automation adoption and innovation (Noble, 1984).

The measure also has important caveats: it captures feasible exposure under stated country contexts, not realised adoption, employment or wage effects, and productivity changes. This distinction is central to recent work stressing how technological capability does not necessarily translate into immediate use (Svanberg et al., 2024; Fleming et al., 2024; McElheran et al., 2024). Indeed, adoption and productivity effects depend on firm scale, complementary investment, management practices, organisational redesign, and local capabilities (Bresnahan et al., 2002; Bloom and Van Reenen, 2007; Cirera et al., 2024), which motivate further research to create novel firm-level measures of automation exposure and adoption. A further limitation is that the O*NET task dictionary is also standardised across country and developed for a high income context. While it enables comparability, it cannot observe every country-, sector-, firm-, or workplace-specific variation in the

[8] Robustness checks preserve the main qualitative pattern. A wider-coverage 90-country specification points to the same broad capability cluster; adding balanced-both to the within-exposed margin decomposition does not change the interpretation; and the earlier permutation-importance ranking gives a similar ordering (Supplementary Figure B.10). A linear Shapley R^2 decomposition shows how explanatory mass is allocated under a linear specification with highly correlated covariates (Supplementary Figure B.11).

task content of work. However, this is a limitation that no existing work has yet addressed, hence motivating new research in creating or augmenting cross-country validated dictionaries of tasks and skills.

Our contribution also points to a broader research agenda. Because the atlas separates exposure by country, sector, channel, and labor margin, it can help link automation measurements to the rekindled debate on industrial policy, helping to understand which complementary investments, capabilities, and conditionalities can turn feasibility into productivity growth while limiting avoidable labour displacement (Cirera and Maloney, 2017; Mazzucato and Rodrik, 2026). Moreover, future work could expand our measurement architecture to the firm level, linking it to observed technology adoption, presence of union representation, and management practices, since industrial relations, managerial practices, and organisational choices are likely key drivers of *which* automation opportunities firms adopt (Bresnahan et al., 2002; Bloom and Van Reenen, 2007; Bloom et al., 2016; Agnolin et al., 2025; Noy and Zhang, 2023).

We argue that our measurement infrastructure can also inform the existing and growing literature on skill training and social protection. A substitutive exposure profile raises different policy concerns from an augmenting or shared human–technology profile. In the first case, the priority is likely to involve income support, job-search assistance, and protection against displacement; in the second, training, certification, and matching policies may help workers redeploy skills toward new tasks. This is especially salient in low- and middle-income countries, where policy design must account for informality, weak insurance systems, and imperfect matching between workers and firms (Nedelkoska and Quintini, 2018; Lassébie and Quintini, 2022; Abebe et al., 2021; Freeman, 2010; Banerjee et al., 2024).

We view automation not as an autonomous and deterministic technological process, but a set of choices over which functions are encoded in machines or digital systems, which forms of discretion remain with workers, and which decisions over how work is performed are transferred to managers, engineers, software, or machines (Noble, 1984). This perspective motivates our focus on tasks, channels, labour margins, and country context and encourages us to investigate in future work what are the economic, social and institutional drivers of automation as an intrinsically human process.

Data availability

The constructed exposure measures are available through the accompanying online atlas at `https://automationatlas.org/`. The public replication package is available at `https://github.com/prashgarg/global-automation-atlas`. It includes the retained task-country labels used in the analyses, country-, occupation-, and industry-level exposure panels, source data for the paper figures and tables, prompt protocols, and data dictionaries. The analysis also uses public source data from O*NET, the World Bank, Penn World Table, Barro–Lee, Worldwide Governance Indicators, CEPII BACI, ILOSTAT, the IMF AI Preparedness Index, Eurostat ICT Usage in Enterprises, and external exposure measures cited in the paper. Source links and their roles in the analysis are documented in the replication package. Underlying public datasets remain subject to their original terms of use.

Code availability

The code used to reproduce the paper figures, tables, and numerical checks is available in the public replication repository at `https://github.com/prashgarg/global-automation-atlas`.

The repository documents the language-model prompts, and the reproducible analysis starts from the retained labels used in the paper.

References

Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., and Quinn, S. (2021). Anonymity or distance? job search and labour market exclusion in a growing african city. *The Review of Economic Studies*, 88(3):1279–1310.

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic Association*, 4(1):37–74.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4B, pages 1043–1171. Elsevier.

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.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Agnolin, P., Anelli, M., Colantone, I., and Stanig, P. (2025). Robots replacing trade unions: Novel data and evidence from western europe. IZA Discussion Paper 17864, Institute of Labor Economics (IZA).

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C. A., and Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science Advances*, 4(7):eaao6030.

Albanesi, S., Dias da Silva, A., Jimeno, J. F., Lamo, A., and Wabitsch, A. (2025). AI and women's employment in europe. *American Economic Review: Papers and Proceedings*, 115:46–50.

Amershi, S., Weld, D. S., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., and Horvitz, E. (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Andrews, D., Nicoletti, G., and Timiliotis, C. (2020). Digital technology diffusion: A matter of capabilities, incentives or both? *European Economic Review*, 128:103513.

Apley, D. W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):1059–1086.

Arntz, M., Gregory, T., and Zierahn, U. (2016). The risk of automation for jobs in oecd countries: A comparative analysis. OECD Social, Employment and Migration Working Papers 189, OECD Publishing, Paris.

Autor, D. and Salomons, A. (2018). Is automation labor share-displacing? productivity growth, employment, and the labor share. *Brookings Papers on Economic Activity*, 2018(Spring):1–87.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the u.s. labor market. *American Economic Review*, 103(5):1553–1597.

Autor, D. H. and Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, 31(S1):S59–S96.

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.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6):775–779.

Banerjee, A., Hanna, R., Olken, B. A., and Sverdlin Lisker, D. (2024). Social protection in the developing world. *Journal of Economic Literature*, 62(4):1349–1421.

Barro, R. J. and Lee, J.-W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

Bloom, N., Sadun, R., and Van Reenen, J. (2016). Management as a technology? NBER Working Paper 22327, National Bureau of Economic Research. Issue Date: June 2016; revised October 2017.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

BLS (2024a). Employment projections, national industry–occupation matrix. Industry–occupation staffing patterns used for occupation-mediated industry weights.

BLS (2024b). Occupational employment and wage statistics (OEWS), national cross-industry estimates. National occupation employment by six-digit SOC, 2024 vintage.

BLS (2024c). Standard occupational classification (soc) to isco-08 crosswalk. Used to map O*NET-SOC occupations to ISCO-08 two-digit groups.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Bresnahan, T. F., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117(1):339–376.

Brussevich, M., Dabla-Norris, E., Kamunge, C., Karnane, P., Khalid, S., and Kochhar, K. (2018). Gender, technology, and the future of work. IMF Staff Discussion Note 2018/007, International Monetary Fund, Washington, DC.

Brussevich, M., Dabla-Norris, E., and Khalid, S. (2019). Is technology widening the gender gap? automation and the future of female employment. IMF Working Paper WP/19/91, International Monetary Fund, Washington, DC.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2025). Generative ai at work. *The Quarterly Journal of Economics*, 140(2):889–942.

Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108:43–47.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114(3):542–551.

Caselli, F. and Coleman, W. J. (2001). Cross-country technology diffusion: The case of computers. *American Economic Review*, 91(2):328–335.

Caunedo, J., Keller, E., and Shin, Y. (2023). Technology and the task content of jobs across the development spectrum. *The World Bank Economic Review*, 37(3):479–493.

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E. J., and Mendes Tavares, M. (2024a). Gen-ai: Artificial intelligence and the future of work. Staff Discussion Note 2024/001, International Monetary Fund.

Cazzaniga, M., Panton, A. J., Li, L., Pizzinelli, C., and Mendes Tavares, M. (2025). A gender lens on labor market exposure to AI. *American Economic Review: Papers and Proceedings*, 115:56–61.

Cazzaniga, M., Pizzinelli, C., Rockall, E. J., and Mendes Tavares, M. (2024b). Exposure to artificial intelligence and occupational mobility: A cross-country analysis. IMF Working Paper 2024/116, International Monetary Fund.

Cirera, X., Comin, D. A., and Cruz, M. (2024). Anatomy of technology and tasks in the establishment. Technical Report 32281, National Bureau of Economic Research.

Cirera, X. and Maloney, W. F. (2017). *The Innovation Paradox: Developing-Country Capabilities and the Unrealized Promise of Technological Catch-Up*. World Bank, Washington, DC.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5):2031–2059.

Comin, D. and Mestieri, M. (2018). If technology has arrived everywhere, why has income diverged? *American Economic Journal: Macroeconomics*, 10(3):137–178.

Das, M. and Hilgenstock, B. (2022). The exposure to routinization: Labor market implications for developed and developing economies. *Structural Change and Economic Dynamics*, 60:99–113.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4):1593–1640.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors*, 59(1):5–27.

Endsley, M. R. and Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):462–492.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182. Version 10.01 used here, available at https://www.rug.nl/ggdc/productivity/pwt/.

Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.

Fetzer, T., Lambert, P. J., Feld, B., and Garg, P. (2024). AI-generated production networks: Measurement and applications to global trade. Discussion Paper 19708, Centre for Economic Policy Research.

Fleming, M., Li, W., and Thompson, N. C. (2024). The last mile problem in AI: Why job automation will be slower than technological progress suggests. Brookings article.

Freeman, R. B. (2010). Labor regulations, unions, and social protection in developing countries: Market distortions or efficient institutions? In Rodrik, D. and Rosenzweig, M., editors, *Handbook of Development Economics*, volume 5, chapter 70, pages 4657–4702. Elsevier.

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.

Gaulier, G. and Zignago, S. (2010). BACI: International trade database at the product-level. the 1994–2007 version. CEPII Working Paper 2010-23, CEPII.

Gmyrek, P., Berg, J., and Bescond, D. (2023). Generative ai and jobs: A global analysis of potential effects on job quantity and quality. ILO Working Paper 96, International Labour Organization.

Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., and Troszyński, M. (2025). Generative ai and jobs: A refined global index of occupational exposure. ILO Working Paper 140, International Labour Organization.

Goos, M. and Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in britain. *Review of Economics and Statistics*, 89(1):118–133.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 104(8):2509–2526.

Graetz, G. and Michaels, G. (2018). Robots at work. *Review of Economics and Statistics*, 100(5):753–768.

Guo, K. and Pan, Y. (2023). Composite adaptation and learning for robot control: A survey. *Annual Reviews in Control*, 55:279–290.

ILO (2012). *International Standard Classification of Occupations (ISCO-08): Structure, group definitions and correspondence tables*. International Labour Organization, Geneva.

ILO (2026). ILOSTAT database. Labour force, employment by occupation (ISCO-08), and employment by economic activity (ISIC Rev. 4).

ILOSTAT (2026a). Labour force statistics (LFS and RURBAN databases). ILOSTAT methods, concepts, definitions, uses, and limitations.

ILOSTAT (2026b). Worker and sector profiles (PROFILES database). ILOSTAT concepts, definitions, interpretation, and limitations.

Lassébie, J. and Quintini, G. (2022). What skills and abilities can automation technologies replicate and what does it mean for workers?: New evidence. OECD Social, Employment and Migration Working Papers 282, OECD Publishing.

Lewandowski, P., Madoń, K., and Park, A. (2025). Workers' exposure to AI across development stages. Discussion Paper 18235, IZA Institute of Labor Economics.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30.

Mazzucato, M. and Rodrik, D. (2026). Industrial policy with conditionalities: A taxonomy and sample cases. *Industrial and Corporate Change*. Article dtaf063.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L. S., and Zolas, N. (2024). AI adoption in america: Who, what, and where. *Journal of Economics & Management Strategy*, 33(2):375–415.

McMillan, M. and Rodrik, D. (2011). Globalization, structural change and productivity growth. In Bacchetta, M. and Jansen, M., editors, *Making Globalization Socially Sustainable*, pages 49–84. World Trade Organization and International Labour Office, Geneva.

Michaels, G., Natraj, A., and Van Reenen, J. (2014). Has ICT polarized skill demand? evidence from eleven countries over twenty-five years. *Review of Economics and Statistics*, 96(1):60–77.

Nedelkoska, L. and Quintini, G. (2018). Automation, skills use and training. OECD Social, Employment and Migration Working Papers No. 202, OECD Publishing, Paris.

Nelson, R. R. and Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *American Economic Review*, 56(1/2):69–75.

Noble, D. F. (1984). *Forces of Production: A Social History of Industrial Automation*. Alfred A. Knopf, New York.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

O*NET Resource Center (2024a). O*net 29.1 database. Archived November 2024 release; standardized task statements, occupation metadata, and skill profiles.

O*NET Resource Center (2024b). O*net online. O*NET-SOC taxonomy and task/tools databases.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3):286–297.

Pizzinelli, C., Panton, A. J., Mendes Tavares, M., Cazzaniga, M., and Li, L. (2023). Labor market exposure to ai: Cross-country differences and distributional implications. IMF Working Paper 2023/216, International Monetary Fund.

Puranam, P. (2021). Human–ai collaborative decision-making as an organization design problem. *Journal of Organization Design*, 10:75–80.

Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1):192–210.

Restrepo, P. (2024). Automation: Theory, evidence, and outlook. *Annual Review of Economics*, 16:1–25.

Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth*, 21(1):1–33.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307.

Svanberg, M., Li, W., Fleming, M., Goehring, B., and Thompson, N. (2024). Beyond AI exposure: Which tasks are cost-effective to automate with computer vision? Working paper, SSRN.

Syed, R., Suriadi, S., Adams, M., Bandara, W., Leemans, S. J. J., Ouyang, C., ter Hofstede, A. H. M., van de Weerd, I., Wynn, M. T., and Reijers, H. A. (2020). Robotic process automation: Contemporary themes and challenges. *Computers in Industry*, 115:103162.

Tomei, P. M. and Klein Teeselink, B. (2026). What jobs can ai learn? measuring exposure by reinforcement learning.

UN Statistics Division (2008). *International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4*. United Nations, Department of Economic and Social Affairs, Statistics Division, New York.

Vaccaro, M., Almaatouq, A., and Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8:2293–2303.

Webb, M. (2020). The impact of artificial intelligence on the labor market. Working paper; SSRN 3482150.

World Bank (2024a). World bank country and lending groups. Analytical classification of economies by income level for fiscal year 2025.

World Bank (2024b). World development indicators. Data retrieved 2026 via the World Bank API.

World Bank (2024c). Worldwide governance indicators.

Yin, M., Vu, H., and Persico, C. (2026). How (un)stable are LLM occupational exposure scores? evidence from multi-model replication. Working Paper 35110, National Bureau of Economic Research.

Supplementary Materials

Global Automation Atlas

Prashant Garg, Tommaso Crosta, and Jasmin Baier

The Supplementary Materials are organised in two notes. Supplementary Note A gives the measurement construction, prompt protocols, and validation checks. Supplementary Note B collects results companions, robustness checks, occupation and industry summaries, and employment-linked extensions.

Supplementary Note A Methodology and Validation Supplement

This note provides additional details related to the validation of the task-country measure used in the main text. It has three parts. First, it defines the analysis objects and the aggregation and linkage rules that turn task labels into country, occupation, and industry summaries. Second, it reports the prompt protocols and output diagnostics needed to interpret the model-produced labels. Third, it reports additional details on the the four validity checks: construct validity, convergent validity across model families, reasoning consistency, and face validity. Results companions are collected in Supplementary Note B.

Supplementary Note A.1 Task aggregation and linkage rules

This subsection details how task labels are normalised, benchmarked, and aggregated into the occupation and industry objects used in the paper.

Task-label normalisation. Each model return is parsed into the fixed schema defined in Section 2.1. Exposure level is stored as a bounded integer on the 0–3 scale; dominant technology channel, labour margin, and dominant AI function are stored as enumerated categorical fields; and AI materiality is stored as a binary flag. For the task-country level dataset, each observation receives one labour-margin label from {substitute, augment, balanced-both, unclear}. The *balanced-both* category is reserved for cases in which the task core is reorganised into a shared human–technology workflow. This convention keeps technology contact, economic exposure, and margin composition separate: Level 1 records assistive contact below the economic threshold, while substitution, augmentation, and balanced-both shares describe only the composition of economically exposed work.

Context variants. The country-conditioned, income-group, and context-free runs use the same task unit, schema, and aggregation rules. Only the context supplied to the classifier changes. The country-conditioned version is the baseline used in the results. The income-group and context-free versions are matched benchmarks used to quantify how much the same task labels change when country information is removed or coarsened to development level.

Occupation linkage details. Let $T(o)$ denote the O*NET tasks attached to parent SOC occupation o and let w_{to} denote task weights within that occupation, normalised so $\sum_{t \in T(o)} w_{to} = 1$. The SOC-level occupation summary is

$$\widehat{E}_o = \sum_{t \in T(o)} w_{to} \, E_t.$$

The ISCO layer is then built by mapping SOC occupations to ISCO groups under bridge shares m_{ok}:

$$\widehat{E}_k^{(w)} = \sum_o m_{ok}^{(w)} \widehat{E}_o, \qquad \widehat{E}_k^{(m)} = \sum_o m_{ok}^{(m)} \widehat{E}_o.$$

We report the task-weighted variant as the baseline and retain the modal variant as a sensitivity check. Analogous summaries are produced for channel, margin, and AI composition using the same bridge weights.

Industry linkage details. For each four-digit ISIC Rev. 4 activity description, embedding similarity generates candidate task links against normalised O*NET task text. Each candidate edge is then passed to a structured LLM voter that decides whether the task is a meaningful component of the activity. The prune pass produces 56,904 completed vote requests over 18,968 candidate edges, with a mean vote-agreement share of 96.6% across repeated prompts. Of the 18,968 candidate edges, 12,294 are retained. The retained graph covers 418 ISIC4 classes and 88 two-digit divisions. We use the same weighted sum as in the occupation summary. Two-digit ISIC figures first summarise each country at the retained four-digit class level and then average equally across constituent four-digit classes within each division.

Employment-weighted reaggregation checks. These checks ask how the constructed exposure measures change when downstream aggregation uses observed employment weights rather than equal task or occupation weights. For the occupation summaries, we reweight the SOC-level measures by U.S. BLS OEWS national employment counts at six-digit SOC (BLS, 2024b) and by the BLS industry–occupation matrix (BLS, 2024a). For the country-level reweighting exercise, we replace the U.S.-based occupation distribution with country-specific ISCO employment shares from ILOSTAT (ILO, 2026; ILOSTAT, 2026a,b). These exercises do not change the underlying task-country labels. They show how much aggregate exposure can move when the same labels are combined with different employment structures, especially at the occupation level (Pizzinelli et al., 2023).

Supplementary Note A.2 Measurement objects, prompt protocols, and output diagnostics

Supplementary Table A.1 lists the main task, country, occupation, industry, and covariate objects used in the paper.

Table A.1: Measurement objects used in the paper.

Object	Identifier	Count	Role
O*NET task universe	standardised task	18,797	Source task unit
Country sample	ISO3 country	124	Country contexts
Country-conditioned task labels	country × task	2,330,776	Main task-country dataset
Benchmark task-context labels	task × context	91,022	Context-free and income-group benchmarks
Country-SOC occupation summaries	country × O*NET-SOC	114,452	Intermediate occupation aggregation
Country-ISCO occupation summaries	country × ISCO-08 major group	1,116	International occupation reporting layer
Retained task-ISIC4 graph	task × ISIC4 edge	12,294	Industry linkage
Country-ISIC4 summaries	country × ISIC4 class	17,112	Industry reporting layer
Country-predictor samples	ISO3 country	68 / 90 / 118	Main, lean, and minimal predictor screens

Notes: Counts refer to the constructed objects used in the paper. The country-conditioned label count is the unique country × task count used for country-level summaries. The larger parsed country × task-code table repeats standardised tasks across O*NET task-code locations. Validation samples are reported in the relevant subsections.

Supplementary Note A.2.1 Prompt protocols and schemas

The tables in this section report the exact schemas used during classification. We provide details on the task-title normalization used for stable matching and display, the automation classifier used for the country-conditioned and benchmark task labels, and the task-to-ISIC4 pruning voter used for the industry linkage. Verbatim prompt text is provided in the supplementary prompt files and replication package; the tables here record the decision object and returned schema.

Task semantic titles. The first stage standardizes O*NET task labels into compact verb-phrase titles that can be compared consistently across later stages. Short verb-phrase titles reduce lexical noise while preserving the granularity needed for retrieval, display, and comparison.

Schema summary: Task semantic titles

Field	Type	Allowed values / enum	Required	Key rule / meaning
short_desc	string	5–10 word observable task descriptor	Yes	Compact semantic title used for retrieval, display, and later prompt conditioning.
is_too_generic	boolean	true | false	Yes	Flags labels that remain broad or residual even after best-effort normalization.

Task automation classifier. This is the core task-labelling stage used for the paper's multidimensional automation-exposure object. The structured JSON separates exposure intensity, mechanism, labour margin, AI materiality, AI function, and short rationales.

Benchmark-context variants. The country-conditioned, income-group, and context-free runs keep the same task unit, evaluation year, and JSON schema. The country-conditioned run uses `COUNTRY CONTEXT: {country_name}` and asks about typical production and institutional settings in the named country around 2026. The income-group benchmark replaces that line with `INCOME GROUP CONTEXT: {income_group}` and asks for a typical country in the named World Bank income group. The context-free benchmark uses `BENCHMARK CONTEXT: Generic ordinary workplace` benchmark and removes geography and income from the prompt. In all three variants, only the context changes.

Schema summary: Task automation classifier

Field	Type	Allowed values / enum	Required	Key rule / meaning
exposure level	integer	0 | 1 | 2 | 3	Yes	Ordered economic-exposure scale from no credible automation margin to extensive economic exposure in the stated context.
dominant channel	string	physical execution; rule-based workflow; planning/control; inference/scoring; information transformation; none	Yes	Mechanism that automates the economically relevant output of the task core.
substitution path	boolean	true | false	Yes	Whether a plausible route substitutes a material part of the task core.
augmentation path	boolean	true | false	Yes	Whether a plausible route materially assists performance while retaining a human role.
margin	string	substitute; augment; both; unclear	Yes	Predominant labour-reallocation route in typical deployment for the stated context.
AI materiality	boolean	true | false	Yes	Whether contemporary learned models are materially necessary to the exposed automation route.
dominant AI function	string	none; state inference; content transformation; recommendation and decision support; adaptive control	Yes	Main role played by learned models when AI materiality is true.
short rationale	string	Max 240 characters	Yes	Brief explanation for the exposure, mechanism, and margin assignment.
substitution summary	string	Max 240 characters	Yes	Main plausible substitution route, if any.
augmentation summary	string	Max 240 characters	Yes	Main plausible augmentation route, if any.

Task-to-ISIC4 pruning voter. Finally, this stage validates candidate links between O*NET tasks and four-digit ISIC Rev. 4 activity descriptions for the industry reporting layer. The task-to-ISIC4 graph starts from embedding-retrieved candidate edges. The pruning prompt applies a conservative activity-component test so the retained graph measures tasks plausibly contained in an industry activity, not loose semantic proximity to an industry label.

Schema summary: Task-to-ISIC4 pruning voter

Field	Type	Allowed values / enum	Required	Key rule / meaning
is_valid	boolean	true | false	Yes	True only when the candidate O*NET task is a materially plausible concrete task component of the stated ISIC4 activity.

Supplementary Note A.2.2 Prompt output diagnostics

This part reports task-automation prompt outputs. The schema is reported in Supplementary Note A.2.1; the table below summarizes what the country-conditioned task automation classifier returns in practice.

Table A.2: Country-conditioned task automation: post-prompt output summary.

Metric	Value
Country-task objects	2,957,459
Unique country×task pairs	2,330,776
Mean valid runs per object	1.00
Economically exposed share (levels 2–3)	41.4%
exposure_level_mode = 0	33.8%
exposure_level_mode = 1	24.8%
exposure_level_mode = 2	34.1%
exposure_level_mode = 3	7.3%
dominant_channel_mode = none	38.5%
dominant_channel_mode = physical_execution	18.5%
dominant_channel_mode = rule_based_workflow	16.1%
dominant_channel_mode = informational_transformation	14.8%
dominant_channel_mode = planning_control	7.0%
dominant_channel_mode = inference_scoring	5.1%
margin_mode = unclear	58.4%
margin_mode = both (balanced-both)	22.6%
margin_mode = substitute	14.9%
margin_mode = augment	4.2%

Notes: Item-level modal outputs after country-task aggregation, computed on the 124-country country-conditioned task dataset. Country-task objects (2,957,459) is the parsed row count; unique country×task pairs (2,330,776) is the deduplicated count used as the denominator throughout the validation checks. Exposure levels 2 and 3 define the paper's economically exposed share; level 1 is assistive-only contact; level 0 is no automation contact. The *unclear* margin row sits at 58.4% because margin is only defined on exposed tasks (41.4%); the substitute/augment/balanced-both share within exposed tasks is recoverable by renormalising the three non-unclear rows.

Country-conditioned task automation outputs.

Supplementary Note A.3 Benchmark conditioning ladder construction

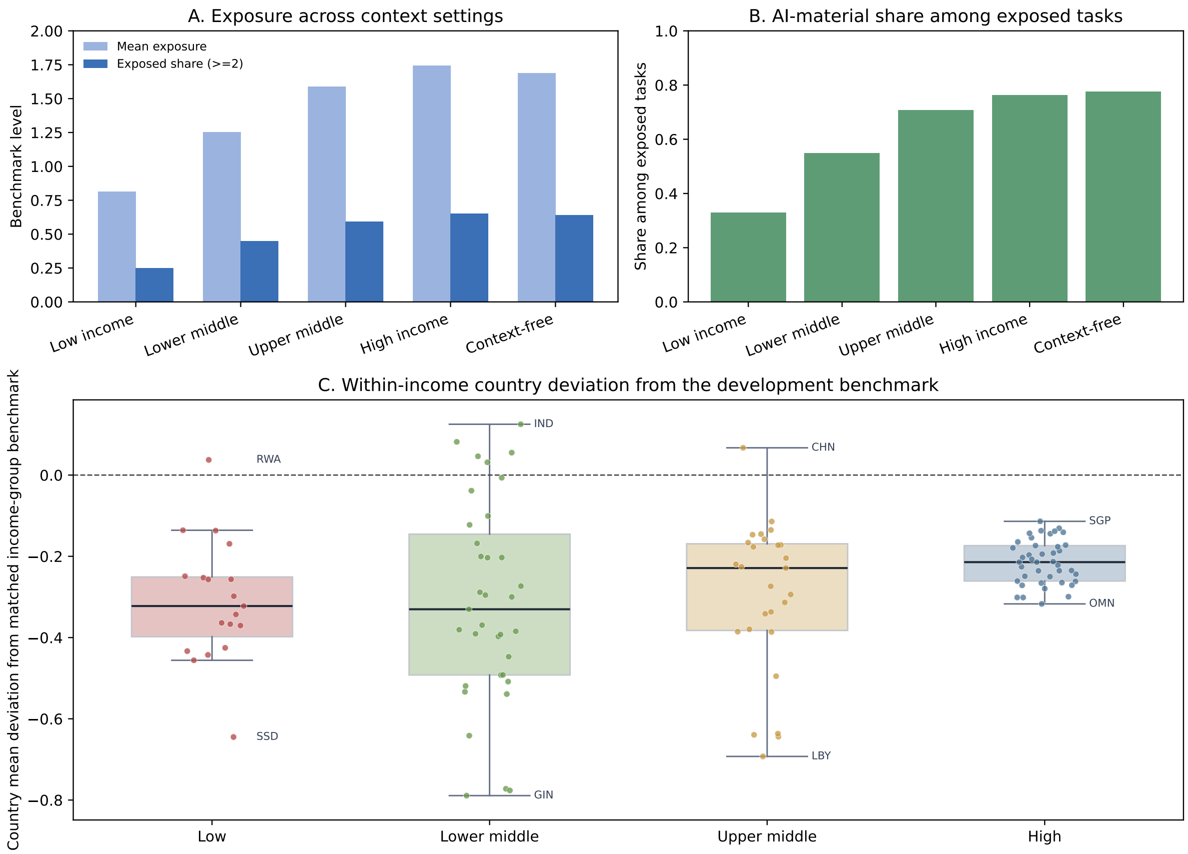
The benchmark ladder separates variation due to the task itself from variation introduced by country context. A context-free label shows how the task is classified when no country information is supplied. Income-group labels show how the same task changes under broad development contexts. Country-conditioned labels show how it changes under the specific production, institutional, and deployment conditions of each country. To make these comparisons interpretable, the labelling pipeline is held fixed across settings: the same task text, schema, response format, model family, and aggregation rules are used throughout. Only the contextual information supplied to the classifier changes.

The ladder has three levels of contextual detail. First, the context-free benchmark labels each task once without naming any country or income group. Second, the income-group benchmarks label each task once under each of the four World Bank income groups: low, lower-middle, upper-middle, and high income. Third, the country-conditioned dataset labels each task separately for each of the 124 countries.

The figure uses the ladder in two ways. Panels (a) and (b) show how headline exposure and AI materiality change as context moves from no country information to broad development context.

Panel (c) isolates variation within income groups. For each country, we subtract the matched income-group benchmark from the country-conditioned task label, task by task, and then average those differences within country. These mean deviations show how much country-level residual exposure remains after broad income-group context has been absorbed.

Figure A.1: Benchmark comparisons across context settings.



Notes: Panel (a) reports mean exposure and exposed share across the context-free and four income-group benchmark prompts. Panel (b) reports AI-material share among exposed tasks for the same benchmark contexts. Panel (c) reports each country's mean task-level deviation from its matched income-group benchmark, using the full country-conditioned dataset.

Supplementary Note A.4 Construct validity findings

Construct validity asks whether our task-level and occupation-level aggregates move with external measures that should capture related constructs. Different instruments measure different sub-constructs of exposure, so the test relies on aligned checks across three surfaces: task- and occupation-level exposure scores from the established literature, a country-level AI readiness composite, and firm-reported AI adoption from official statistics.

We retain four external sources that together capture different notions of automation or AI relevance at different levels and vintages: Frey and Osborne (2017), Felten et al. (2021), Webb (2020), and Eloundou et al. (2024). Supplementary Table A.3 makes explicit which external variable is compared with which paper variable. Supplementary Tables A.4 and A.5 report the task- and occupation-level summaries.

Eloundou et al. report task-level exposure ratings for GPT-style systems and occupation-level aggregates under alpha, beta, and gamma definitions. We use their task-level ratings for the task-level comparison. At the occupation level, we treat the gamma series as a comparator for foundation-model-era AI systems because it is the broadest reported Eloundou aggregate and allows exposure through complementary foundation-model-powered software. In our schema, the closest construct-aligned variable is the foundation-model-like channel share: exposed work whose dominant route is inference/scoring or informational transformation. This is narrower than overall AI-material exposure, which also includes other AI functions such as adaptive control or recommendation. We therefore report AI-material exposure as a broad AI comparison, while reading the foundation-model-like channel comparison as the closest match to Eloundou gamma.

Aggregation scope changes the comparison, so we use four summaries where useful: (i) US-only, matching the native context of several occupation-level comparators; (ii) GDP-weighted, emphasizing output-weighted production environments; (iii) population-weighted, giving more weight to where workers live and therefore central for the development interpretation; and (iv) unweighted, treating national settings symmetrically.

Table A.3: External comparator variables and matched paper variables.

Source	Other paper variable	Our variable	Rows
Frey–Osborne	Occupation computerisation probability	Output-weighted mean occupation exposure	582
Felten AIOE	AI Occupational Exposure index	Output-weighted mean occupation exposure; AI-material share in channel-aligned checks	665
Webb AI	AI patent-text exposure score	Output-weighted mean occupation exposure; AI-material share in channel-aligned checks	762
Webb Software	Software patent-text exposure score	Software-like channel share	762
Webb Robotics	Robotics patent-text exposure score	Physical-execution channel share	762
Eloundou GPT-4 task ratings	Task-level GPT-4 exposure rating, Eloundou rubric	Output-weighted task exposure	23,743
Eloundou human task ratings	Task-level human exposure rating, Eloundou rubric	Output-weighted task exposure	116
Eloundou GPT-4 gamma	Occupation-level gamma score from GPT-4 task ratings	Output-weighted mean occupation exposure; AI-material share as a broad AI check; foundation-model-like share for channel checks	923
Eloundou human gamma	Occupation-level gamma score from human task ratings	Output-weighted mean occupation exposure; AI-material share as a broad AI check; foundation-model-like share for channel checks	923

Notes: Rows list external benchmark variables used in the appendix comparator suite and the paper variable matched to each one. The comparison unit differs across sources.

Table A.4: Task-level comparison of the reported output-weighted global summary against external task-native benchmarks.

Benchmark	Overlap	Pearson	Spearman	Mean abs movement (pp)
GPTs GPT-4 task (Eloundou)	23,743	0.533	0.575	19.1
GPTs human task (Eloundou)	116	0.764	0.832	0.1

Notes: Rows compare paper task-level measures with Eloundou task-level GPT-4 and human-labelled exposure scores. Paper measures are computed from the reported output-weighted global task summary. Pearson and Spearman correlations use the matched task sample shown in each row.

Table A.5: Occupation-level comparison of the reported output-weighted global summary against external occupation benchmarks.

Benchmark	Overlap	Pearson	Spearman	Mean abs movement (pp)
GPTs GPT-4 gamma (Eloundou)	923	0.406	0.405	25.3
GPTs human gamma (Eloundou)	923	0.387	0.359	26.1
Felten AIOE	665	0.313	0.261	20.6
Frey-Osborne	582	0.292	0.358	16.6
Webb AI	762	0.138	0.178	24.7

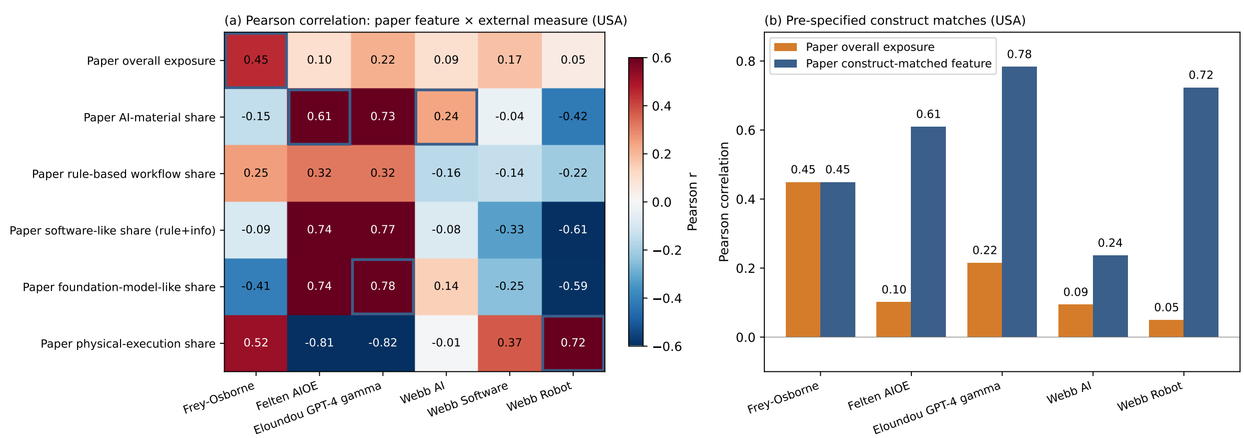
Notes: Rows compare the output-weighted atlas mean occupation exposure with external occupation benchmarks; output weights are country GDP shares. Eloundou scores are aggregated to occupations using the GPT-4 and human gamma definitions in the source; other measures are occupation-native. Mean absolute movement is reported in percentile points within the matched occupation universe.

The test we use at each surface is *sub-construct alignment*. Some external measures map cleanly to one object in our schema: AI-focused measures to AI-material share, foundation-model measures to inference and informational-transformation channels, and robotics measures to physical execution. Other measures combine several mechanisms. Frey–Osborne is a computerisation probability, so its matched comparison is overall exposure. Webb's software-patent score remains in the full matrix as a mixed construct, since software patents can describe workflow software, industrial-control software, and software embedded in physical automation. Our measure differs from every external measure on at least one construct margin: it is country-conditioned, it distinguishes substitution from augmentation, it separates AI materiality from legacy automation in the same schema, and it uses a four-level ordinal scale. The comparators also differ in native unit: Eloundou's scores are task-native and match our task labels directly; Frey–Osborne, Webb, and Felten are native at the occupation level, so their correlations are reported for the occupation summaries. We treat Eloundou et al. as an important foundation-model-era exposure comparator. Recent evidence shows that LLM-generated occupational exposure scores can shift substantially when the same scoring scheme is applied by different rating models (Yin et al., 2026). For this reason, the construct-validity exercise focuses on whether agreement is strongest along matched dimensions, without requiring exact agreement with any single external score.

Task- and occupation-level exposure scores. At the task level, Eloundou's task-native scores are the only like-for-like external comparator. Against our reported output-weighted global task summary, Eloundou's GPT-4 task scores correlate at Pearson 0.533, Spearman 0.575 across 23,743 comparable tasks, and the Eloundou human-labelled subset correlates at Pearson 0.764, Spearman 0.832 across the 116 tasks their human labelling covers. These correlations are high for independently constructed foundation-model exposure instruments and are consistent with the two measures capturing related task-level constructs. At the occupation level our reported output-weighted summary correlates with Eloundou et al. (2024) GPT-4 gamma scores at Pearson 0.406, Spearman 0.405 across 923 occupations; with Eloundou human-labelled gamma at Pearson 0.387, Spearman 0.359; with Felten et al. (2021) AIOE at Pearson 0.313, Spearman 0.261; and with Frey and Osborne (2017) computerisation probability at Pearson 0.292, Spearman 0.358. Webb (2020)'s AI sub-score correlates at Pearson 0.138, the lowest of the group. Agreement is strongest with the Eloundou task-level comparator, moderate with the capability-based occupation measures (Felten, Frey–Osborne), and weakest with Webb's patent-text AI measure. This ordering is expected: Webb AI captures patented AI applications, whereas much of the current foundation-model deployment margin runs through general-purpose models, APIs, and open-source systems rather than firm-held AI patents. We retain Webb AI in the table for completeness. Webb robotics gives the cleanest Webb channel comparison; Webb software is reported as a mixed patent comparator.

The matched comparisons are sharper than overall exposure where the external construct is narrow. Figure A.2 panel (a) reports the full feature-by-comparator Pearson matrix on the US occupation universe; panel (b) contrasts overall exposure with the pre-specified feature for each clean match. Frey–Osborne is broad, so its matched feature is overall exposure (Pearson 0.45). For the narrower comparators, the matched feature is much closer than overall exposure: foundation-model-like share correlates with Eloundou GPT-4 gamma at Pearson 0.78, compared to 0.22 for overall exposure; AI-material share correlates with Felten AIOE at 0.61, compared to 0.10; and physical-execution share correlates with Webb robotics at 0.72, compared to 0.05. The off-diagonal cells clarify the mixed cases. Webb software correlates negatively with rule-based workflow share (−0.14) and software-like share (−0.33), but positively with physical-execution share (0.37), consistent with software patents often describing industrial-control and process-automation systems. The Webb robotics 0.72 correlation is the strongest clean channel correspondence in the matrix: Webb robotics measures patented industrial and service robotics, and our physical-execution channel captures the same underlying construct in its task-level labels.

Figure A.2: Construct-matched validity: paper sub-constructs against external sub-scores.



Notes: Correlations are computed across 923 US six-digit O*NET-SOC occupations. Paper measures come from the US context-conditioned task labels; external measures are Frey–Osborne computerisation probability, Felten AIOE, Eloundou GPT-4 gamma, and Webb's AI, software, and robotics sub-scores. Panel (a) reports the Pearson-correlation matrix. Outlined cells mark the pre-specified construct matches shown in panel (b). Webb Software is unoutlined because its patent construct mixes software-only and software-enabled physical automation.

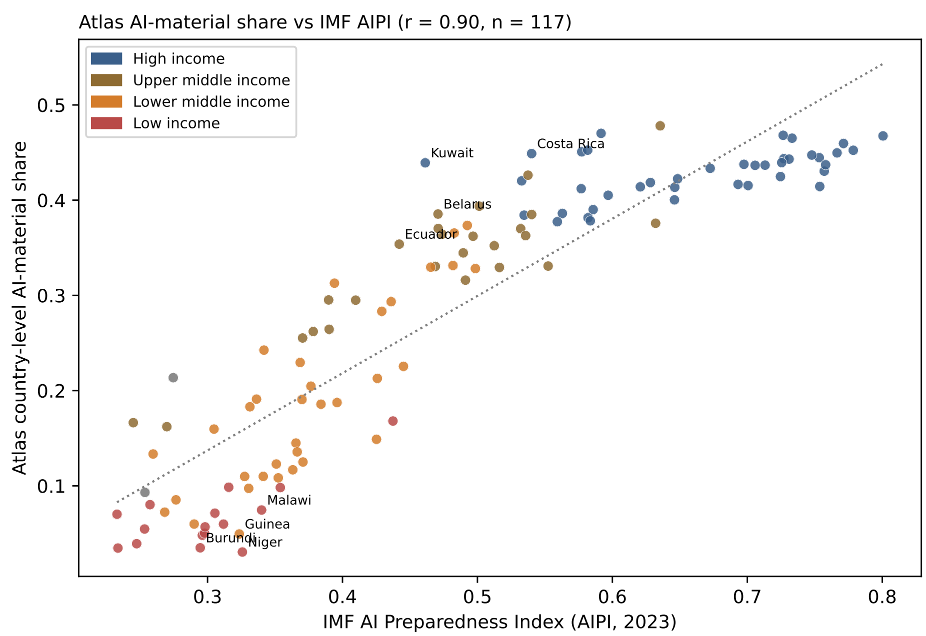
Country-level AI readiness. A further external anchor asks whether our country-level AI-material pattern agrees with an independently constructed country-level readiness measure. The

IMF AI Preparedness Index (AIPI) reported by Cazzaniga et al. (2024a) is the closest comparator: it scores economies on digital infrastructure, human capital and labour-market policies, innovation and economic integration, and regulation and ethics. We compare the AIPI composite and its sub-components with country-level AI-material share, foundation-model-like share, and overall mean exposure across the countries with both paper and IMF coverage. Figure A.3 reports the main comparison; Table A.6 reports the raw, partial, within-income, rank, and leave-one-out checks.

The raw relationship is strong. AI-material share correlates with the AIPI composite at Pearson 0.90 (Spearman 0.93); overall mean exposure correlates at Pearson 0.88 (Spearman 0.95). The comparison is deliberately demanding because both variables are closely related to development: AI-material share correlates with log GDP per capita at Pearson 0.92, and AIPI correlates with log GDP per capita at Pearson 0.91. We therefore report the log-GDP-residual association as a stricter check on whether the two measures align beyond their common development gradient. After residualising both variables on log GDP per capita, the AI-material–AIPI partial Pearson remains positive at 0.42.

Within-income comparisons give the same reading. AI-material share still correlates with AIPI at Pearson 0.72 in low-income countries, 0.84 in lower-middle-income countries, and 0.90 in upper-middle-income countries; the high-income correlation is lower (0.50), where the AIPI approaches its ceiling. The residual disagreements are interpretable: our task-level country conditioning gives relatively more AI-material exposure to economies with specific sectoral technology deployment, and less to economies where composite readiness scores run ahead of task-level deployment conditions. The AIPI check therefore supports the country-level ordering while making clear that the raw 0.90 correlation partly reflects shared development-stage information.

Figure A.3: Country-level AI-material share against the IMF AI Preparedness Index (AIPI).



Notes: Points are countries with coverage in both datasets. The x-axis is the IMF AIPI composite for 2023; the y-axis is country-level AI-material share. The dotted line is a linear fit. Pearson correlation is 0.90 and Spearman correlation is 0.93. Colours mark World Bank income groups; Table A.6 reports robustness checks.

Table A.6 compiles the raw and partial Pearson correlations, the within-income-group Pearson correlations, the rank-robustness check, and a leave-one-out stability check into one reference panel.

Leave-one-out robustness is tight: dropping any single country of the 117 changes the Pearson within $[0.897, 0.905]$ (SD 0.001), and dropping the US and China jointly moves the coefficient from 0.899 to 0.898. No single country shifts the correlation materially.

Table A.6: Atlas–AIPI cross-country correlation checks.

Panel A. Raw and log-GDP-residual Pearson correlations, atlas AI-material share × IMF AIPI series.

IMF series	Raw r	Residual r	Raw-to-residual change	n	Interpretation
AIPI composite	0.899	0.417	54%	117	Positive residual association beyond the common GDP gradient.
Digital Infrastructure (DI)	0.914	0.498	46%	122	Largest residual association.
Human Capital & LM Policies (HCLMP)	0.858	0.378	56%	121	Similar profile to composite.
Innovation & Economic Integration (IEI)	0.750	0.070	91%	118	Closely tied to the GDP gradient in this sample.
Regulation & Ethics (RE)	0.835	0.319	62%	122	Positive residual association after GDP adjustment.

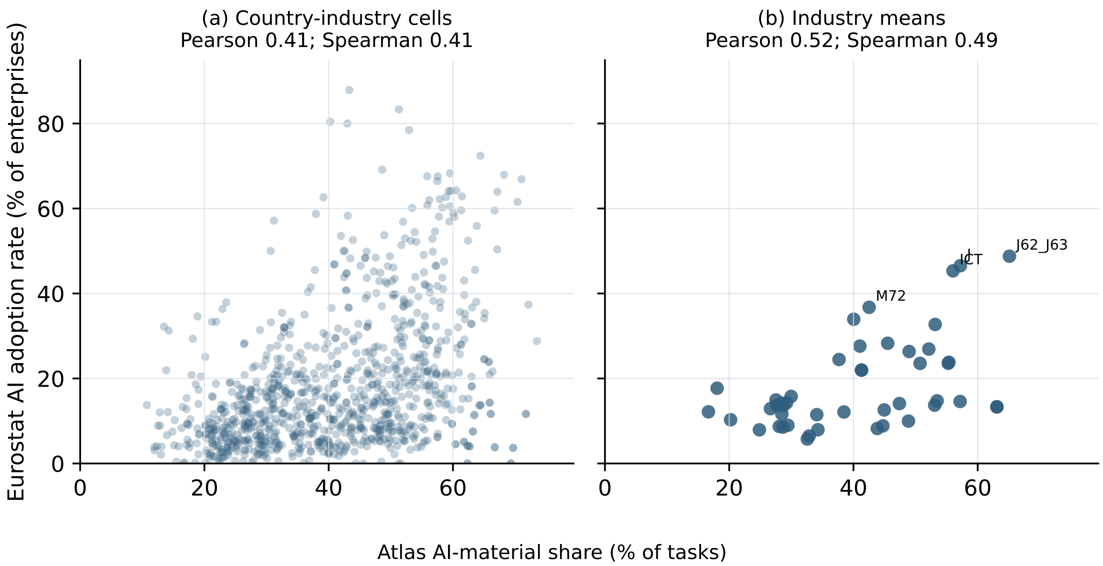
Panel B. Within-income-group Pearson correlations, atlas features × AIPI composite.

Income group	n	AI-material vs AIPI	FM-like vs AIPI	Interpretation
Low income	19	0.718	0.710	Strong residual correspondence despite compressed range.
Lower middle income	35	0.839	0.826	Strongest within-group agreement.
Upper middle income	25	0.897	0.814	Strongest within-group agreement.
High income	41	0.504	0.038	AIPI approaches ceiling; atlas still varies.

Notes: Rows report alternative correlations between country-level AI-material share and the IMF AIPI composite. FM-like denotes the foundation-model-like channel bundle. Residual correlations are computed by residualising both series on the named controls and then correlating the residuals. Within-income rows remove World Bank income-group means. Leave-one-out rows report the range after dropping one country at a time.

Firm-level AI adoption. The construct-validity correlations above use comparators that are language-model-produced (Eloundou) or derived from occupation-text or patent-text procedures (Frey–Osborne, Felten, Webb). Eurostat firm reports provide a different check: whether industries with higher AI-material exposure also report more AI use. This comparison helps separate agreement with other model- or text-based measures from agreement with observed firm behaviour. We use the Eurostat ICT Usage in Enterprises AI module, which asks a representative sample of enterprises with 10 or more employees across 36 European countries whether they use any AI technology, broken out by reported NACE cells. We extract the headline indicator (the share of enterprises reporting any AI use) for 2024 (with 2023 fallback), map NACE codes to ISIC-2 divisions using the standard harmonisation, and correlate against the country × ISIC-2 industry-level features constructed from the bottom-up retained task-to-ISIC-4 graph and the task-country level dataset. Figure A.4 reports both the pooled country–NACE cells and the NACE-cell means on common axes. The relationship has the expected sign. AI-material share correlates with observed AI adoption at Pearson 0.41 pooled across 963 reported country–NACE cells (23 countries, 46 NACE cells), and at Pearson 0.52 across reported NACE-cell means. The pooled rank correlation is also positive (Spearman 0.41); the NACE-cell mean rank correlation is 0.49. Within-country correlations (AI-material ranking versus AI-adoption ranking across reported NACE cells within each country) have mean Pearson 0.42, median 0.43, with 23 countries contributing sufficient overlap for the test. Dropping one country at a time leaves the pooled Pearson in $[0.40, 0.42]$, and dropping one NACE cell at a time leaves it in $[0.36, 0.44]$. Physical-execution share is negatively correlated with AI adoption at Pearson -0.52 across reported NACE-cell means, consistent with physical-heavy industries being less AI-adopting and more robotics-oriented in this survey frame. Overall exposure, which combines all five channels, correlates only at Pearson 0.13 pooled. The comparison therefore supports the AI-material distinction: the AI-specific component of exposure lines up more closely with reported AI use than the broad exposure aggregate. The relationship should still be read as an industry-level cross-check, since realised adoption also depends on firm scale, deployment costs, managerial capacity, regulation, and survey timing.[9]

Figure A.4: Industry AI-material share against observed Eurostat firm-level AI adoption.



Notes: Both panels use common axes. Panel (a) plots 963 matched country × reported NACE cells; panel (b) plots 46 reported NACE-cell means. Reported NACE cells include Eurostat aggregates and sub-aggregates, so they should be read as survey reporting cells rather than mutually exclusive industries. The vertical axis is the Eurostat share of enterprises using any AI technology, using 2024 values with 2023 fallback. AI-material share correlates positively with firm-reported AI adoption in both panels.

Construct validity here rests on agreement across instruments with different constructs and vintages. The evidence is strongest in two places: (i) our measure agrees with the other task-level exposure measure for foundation-model-era AI systems (Eloundou task Pearson 0.533; foundation-model-aligned channel share 0.78), with the AI-specific occupation measure (Felten AIOE aligned to AI-material share at 0.61), and with the robotics-specific patent measure (Webb robotics aligned to physical-execution share at 0.72); and (ii) the observational cross-checks point in the expected direction, including Eurostat firm-reported AI adoption. These checks support external alignment without treating any single comparator as a ground truth. The country-conditioning check that follows asks whether those labels add information beyond US-centric or context-free baselines.

[9] A second, coarser observational anchor is the Stanford AI Index / McKinsey compiled AI-adoption-by-sector dataset, which pools McKinsey State of AI, IBM AI Adoption Index, PwC, and Stanford AI Index figures into sector-level annual adoption rates across 17 industry groupings. Mapping those groupings to ISIC-2 divisions, 2024–2025 sector adoption correlates with AI-material share at Pearson 0.34 across 15 matched sectors. The weaker correlation relative to the Eurostat reading reflects the much coarser industry aggregation (a single "Manufacturing" bucket spans 24 ISIC-2 divisions) and the mixed consulting-survey methodology. The positive correlation across a different reporting mode and industry taxonomy provides a cross-source robustness check on the sign of the relationship.

Supplementary Note A.5 Internal validity and model-robustness checks

This subsection reports the checks that operate directly on the model-produced labels and prompt outputs. It begins with cross-model convergence, then turns to rationale and prompt consistency, and closes with direct inspection of label distributions and anchor occupations.

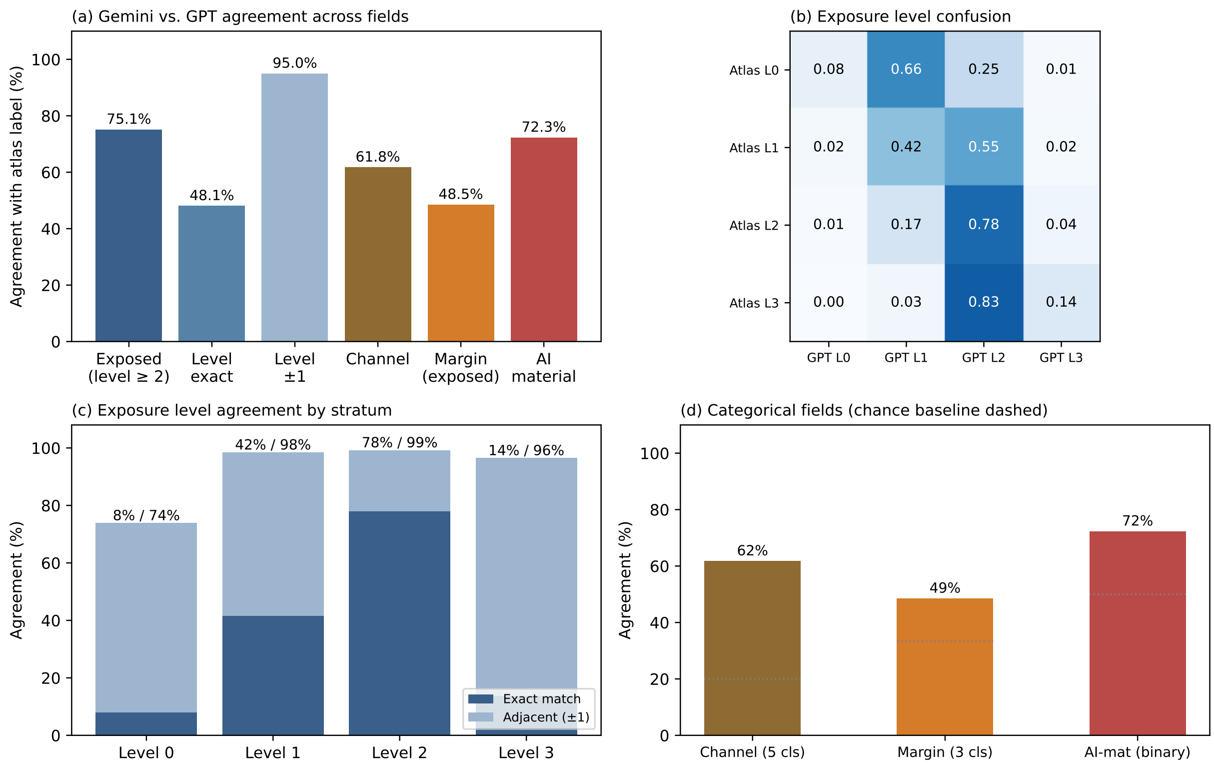
Supplementary Note A.5.1 Convergent validity across model families

Convergent validity asks how far the measurement travels across model families when the schema and prompt are held fixed. We re-label the full context-free task universe with OpenAI's `gpt-5.4-mini` under the same context-free system prompt and JSON response schema, and compare those labels with the main run. Figure A.5 reports results for 18,169 tasks with valid cross-model predictions, a 99.96% parse success rate. The check speaks most clearly to the ordinal structure of the exposure scale: 95.0% of tasks receive a cross-model prediction within one level of the original label. The exact exposed-versus-not-exposed cutoff is more sensitive to the rating model: binary exposed agreement is 75.1%, and exact four-level agreement is 48.1%.

The disagreement pattern is systematic. GPT compresses labels toward the middle of the four-level scale, rating many original level-0 tasks as level 1 or 2 and many original level-3 tasks as level 2. This compression lowers agreement at the exposed threshold while preserving most adjacent-level ordering. Agreement on the other fields is lower than for the ordinal exposure scale but remains above random baselines: dominant channel matches exactly on 61.8% of tasks, AI materiality on 72.3%, and labour margin among exposed tasks on 48.5%.

Taken together, the check supports ordinal comparisons most clearly. The level-2 cutoff is the harder object: it still agrees in three quarters of cases, but it carries more model-family sensitivity than adjacent ordering. We therefore report both numbers and treat the cutoff as the demanding part of this validation exercise. The present check uses context-free tasks; applying the same design to country-conditioned labels and further model families is the natural extension.

Figure A.5: Cross-model convergent validity: main labels vs. `gpt-5.4-mini` over the full context-free task universe.



Notes: Both models receive the original context-free system prompt, task text, and JSON response schema. Valid cross-model predictions cover 99.96% of the context-free universe. Panel (a) reports agreement by label field; panel (b) reports the exposure-level confusion matrix; panel (c) reports exposure agreement by original level; panel (d) compares categorical-field agreement with random baselines.

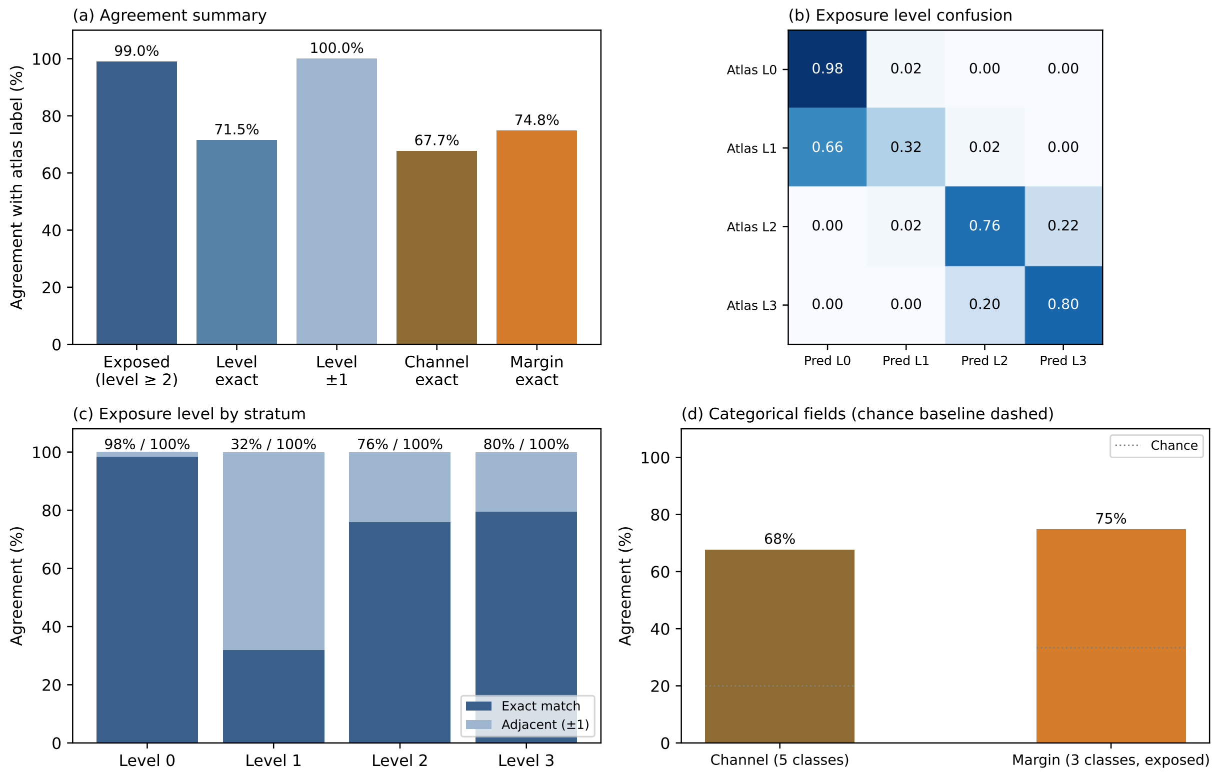
Supplementary Note A.5.2 Reasoning consistency

Reasoning consistency asks whether the model's stated rationale supports the label attached to it, and whether superficial prompt rewording changes the classification. This is an internal-coherence check. A rationale is informative only if it contains enough information to recover the label, and a prompt is more credible if superficial wording changes leave the classification largely unchanged. The evidence is strongest for the headline exposed-versus-not-exposed distinction. Agreement is lower for channel and labour-margin fields, as expected for multi-class labels that require finer distinctions.

Rationale-to-label predictability. We draw a stratified sample of 1,000 country-task observations, with 250 observations from each exposure level. For each observation, an independent model family receives the task description and the original rationale, with the original label withheld, and predicts the exposure level, dominant technology channel, and labour margin under the paper's schema. The independent model recovers the exposed-versus-not-exposed label on 99.0% of observations; only 10 of 1,000 predictions cross the exposed threshold. Exposure level is within one step of the original label for all observations, with exact-level agreement at 71.5%. Agreement is lower, but still well above random baselines, for dominant channel (67.7% exact agreement across five classes) and labour margin among exposed tasks (74.8% exact agreement across three classes).

Disagreements concentrate at adjacent exposure levels, especially around the level-0/level-1 and level-2/level-3 boundaries. The check therefore supports the internal coherence of the headline exposure label, while showing that fine-grained fields should be read with more caution.

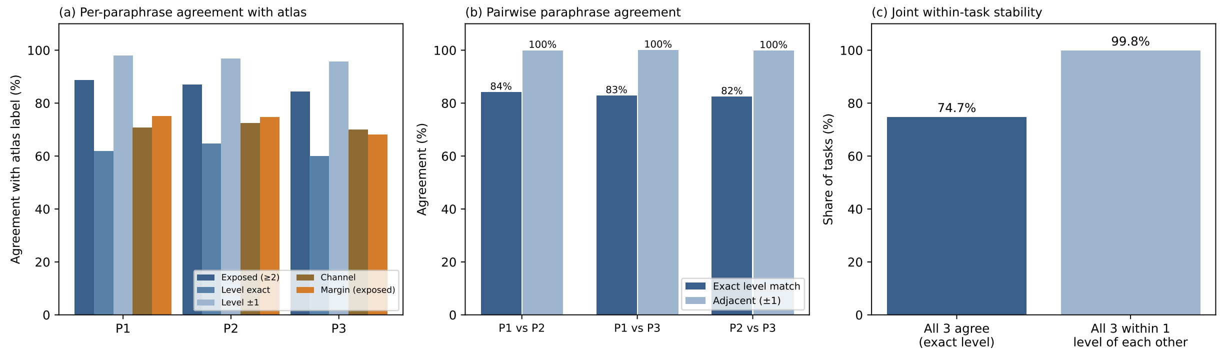
Figure A.6: Rationale-to-label predictability under an independent model family.



Notes: The independent model is `gpt-5.4-mini`. It receives the task description and original rationale, with the original label withheld, for 1,000 country-task observations stratified equally across exposure levels. Panel (a) reports agreement by label field; panel (b) reports the exposure-level confusion matrix; panel (c) reports exposure agreement by original level; panel (d) compares channel and margin agreement with random baselines.

Prompt-paraphrase stability. We next test whether superficial changes to the prompt wording change the labels. An independent model (GPT-5.4, used only for paraphrasing) generates three reworded versions of the context-free system prompt. The paraphrases preserve all schema keys, value spaces, numeric thresholds, and consistency rules, while changing wording, sentence structure, and section order. We apply the three paraphrases to a stratified sample of 1,000 tasks using the same `gemini-3.1-flash-lite` model, response schema, and temperature settings as the original labelling run. Agreement with the original exposed-versus-not-exposed label is 88.7%, 87.0%, and 84.3% across the three paraphrases; exposure level is within one step of the original label for 96%–98% of tasks. Across the three paraphrases, 99.8% of tasks sit within one exposure level of each other. The remaining disagreements mostly occur at adjacent exposure levels and rarely cross the exposed threshold. Prompt wording therefore has limited practical leverage over the headline exposure classification, though exact levels and compositional fields remain more sensitive.

Figure A.7: Paraphrase stability under three rewordings of the system prompt.



Notes: Three paraphrases of the context-free system prompt are applied to 1,000 tasks stratified equally across exposure levels. The model, response schema, and temperature settings match the original run. Panel (a) compares each paraphrase with the original labels; panel (b) reports pairwise exposure-level agreement across paraphrases; panel (c) reports joint within-task stability across all three paraphrases.

Supplementary Note A.5.3 Face validity

Face validity asks whether the task-level labels and accompanying rationales make sense under direct inspection, and whether their distribution across the task-country level dataset matches the measurement objects defined in Section 2.1. The check targets structurally implausible output that direct inspection would catch. We use four diagnostics. First, label-field distributions check for degenerate or mechanically bunched outputs (Figure A.9, Figure A.10). Second, a rationale–label consistency check scans every observation in the task-country level dataset (Table A.8). Third, a rationale-divergence test asks whether country conditioning changes the substantive content of the rationale (Figure A.8, Table A.9). Fourth, an anchor-occupation table compares our measure with occupations where the existing literature has strong priors (Supplementary Table A.7). The four diagnostics do not identify a face-validity problem.

Table A.7: Exposure for anchor occupations where the existing literature has strong priors.

Occupation	Prior direction	Mean exposure	High income	Low income	Sub. share	Aug. share
Childcare workers	Very low	0.42	0.64	0.10	0.04	0.02
Elementary school teachers	Very low	0.68	0.93	0.25	0.04	0.03
Surgeons	Very low	0.72	1.04	0.24	0.00	0.11
Waiters and waitresses	Medium	0.94	1.40	0.32	0.19	0.00
Farm labourers (crops)	Medium	1.04	1.59	0.33	0.13	0.01
Heavy-truck drivers	Medium	1.36	1.86	0.62	0.30	0.01
Software developers	High	1.26	1.62	0.66	0.04	0.12
Interpreters and translators	High	1.50	1.70	1.07	0.13	0.03
Cashiers	High	1.72	2.12	0.97	0.51	0.01
Accountants and auditors	High	1.92	2.32	1.22	0.30	0.02
Bookkeeping and audit clerks	High	2.54	2.91	1.85	0.85	0.00

Notes: Mean exposure is the cross-country unweighted occupation-level mean on the 0–3 scale. High- and low-income columns report the corresponding World Bank income-group means. Substitution and augmentation shares are task-weighted occupation means averaged across countries. The prior column summarises expectations from the existing exposure literature (Frey and Osborne, 2017; Webb, 2020; Felten et al., 2021; Eloundou et al., 2024); rows are grouped by prior direction and sorted by mean exposure within group.

Distributional check. We first inspect the marginal distribution of each task-country label field. The exposure labels span all four levels without degeneracy, and the dominant-channel field is not concentrated in a single technology category. Among exposed tasks, the labour-margin split is mostly balanced-both or substitution-only, with a smaller augmentation-only share. The AI-materiality flag is true for 27.8% of observations, consistent with our interpretation that most measured exposure is carried by non-AI channels.

Income-group check. We then repeat the distributional check by World Bank income group. The label distributions move in the expected direction: level-0 and no-channel shares fall with income, while AI-materiality rises from 6% in low-income countries to 43% in high-income countries. These gradients match the cross-country patterns reported in Sections 3.1–3.3, so the raw label distributions are coherent with the paper's main descriptive facts.

Rationale–label consistency. We next ask whether the rationale text ever directly contradicts the label attached to it. This is the most mechanical internal-coherence check in the appendix because it is applied to the full task-country dataset, covering all 2,330,776 unique task-country rationales. We use five targeted rules that correspond to sharp schema conflicts: a level-3 rationale denying automation is possible; a level-0 rationale describing standard automation as already available; an *augment* rationale describing replacement; a *substitute* rationale describing assistive-only technology; and a not-AI-material rationale invoking AI, LLM, or machine-learning capabilities. A sentence-level negation filter suppresses matches in phrases such as "not widely deployed" or "lacks a credible AI-based substitute."

The screen flags 2,654 observations, or 0.11% of the task-country dataset (Table A.8). We read this as evidence against widespread mechanical inconsistency between labels and rationales. The number is not an adjudicated error rate: keyword rules can miss subtle contradictions and can also flag benign wording. Its value is that the most direct label-rationale conflicts are rare at full-dataset scale. The rationale-to-label predictability exercise below then tests the same issue with an independent model rather than keyword rules.

Table A.8: Rationale–label consistency check over the task-country level dataset.

	Eligible rows	Keyword-flagged	Share
Rule: level-3 rationale denies automation	175,802	19	0.01%
Rule: level-0 rationale describes automation	804,336	549	0.07%
Rule: augment rationale describes replacement	95,375	1	0.00%
Rule: substitute rationale describes assistive-only	357,000	0	0.00%
Rule: not-AI-material rationale invokes AI/LLM/ML	1,683,883	2,099	0.12%
Unique observations flagged by any rule	2,330,776	2,654	0.11%

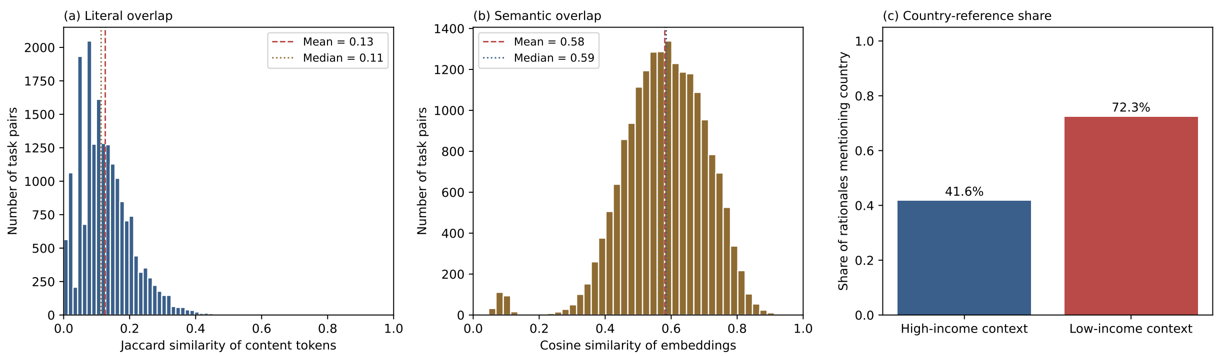
Notes: Each row scans eligible rationales for phrases that would indicate a potential contradiction with the attached label. Eligible rows are observations satisfying the relevant label condition. A sentence-level negation filter suppresses matches in negating clauses, so flagged rows should be read as potential inconsistencies rather than adjudicated errors.

Country-conditioning divergence. This check asks whether country conditioning changes the substance of the classifier's reasoning. The concern is that the model could produce nearly the same rationale for a task regardless of country, with only the country name or a generic constraint added. To test this, we compare rationales for the same O*NET task under one high-income and one low-income country prompt.

We use two measures because the two rationales can differ in wording while still making the same argument. First, we compute Jaccard overlap on stopword-filtered content tokens, which measures direct word reuse. Second, we embed each rationale using OpenAI `text-embedding-3-large` and compute cosine similarity, which measures semantic similarity. Low Jaccard overlap with moderate cosine similarity means the rationales discuss the same task but use different reasoning.

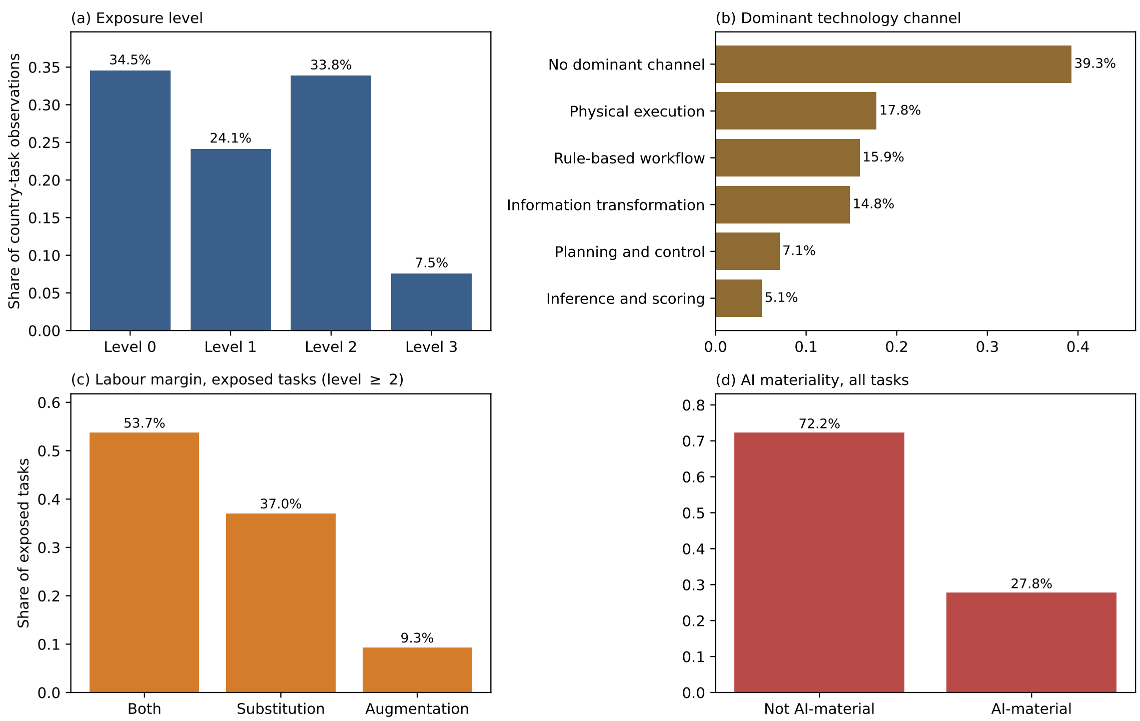
The results are consistent with country conditioning changing the substance of the explanation. Most pairs have low Jaccard overlap and moderate cosine similarity: they remain on topic, but describe different production environments, skill levels, and adoption constraints. Only 0.4% of pairs fall in the high-Jaccard / high-cosine region, where country conditioning would look close to near-verbatim reuse. Low-income rationales also name the country more often than high-income rationales, consistent with deployment constraints being more explicit in lower-capability settings.

Figure A.8: Country conditioning changes the substance of rationales while keeping them on-topic.



Notes: Each task contributes one randomly drawn high-income rationale and one randomly drawn low-income rationale. Panel (a) reports Jaccard similarity on stopword-filtered content tokens; panel (b) reports cosine similarity using OpenAI `text-embedding-3-large` embeddings; panel (c) reports whether the rationale names the prompted country. Low word overlap with moderate semantic similarity indicates that rationales remain on-task while changing their country-level reasoning.

Figure A.9: Distributional check of the task-level labels.



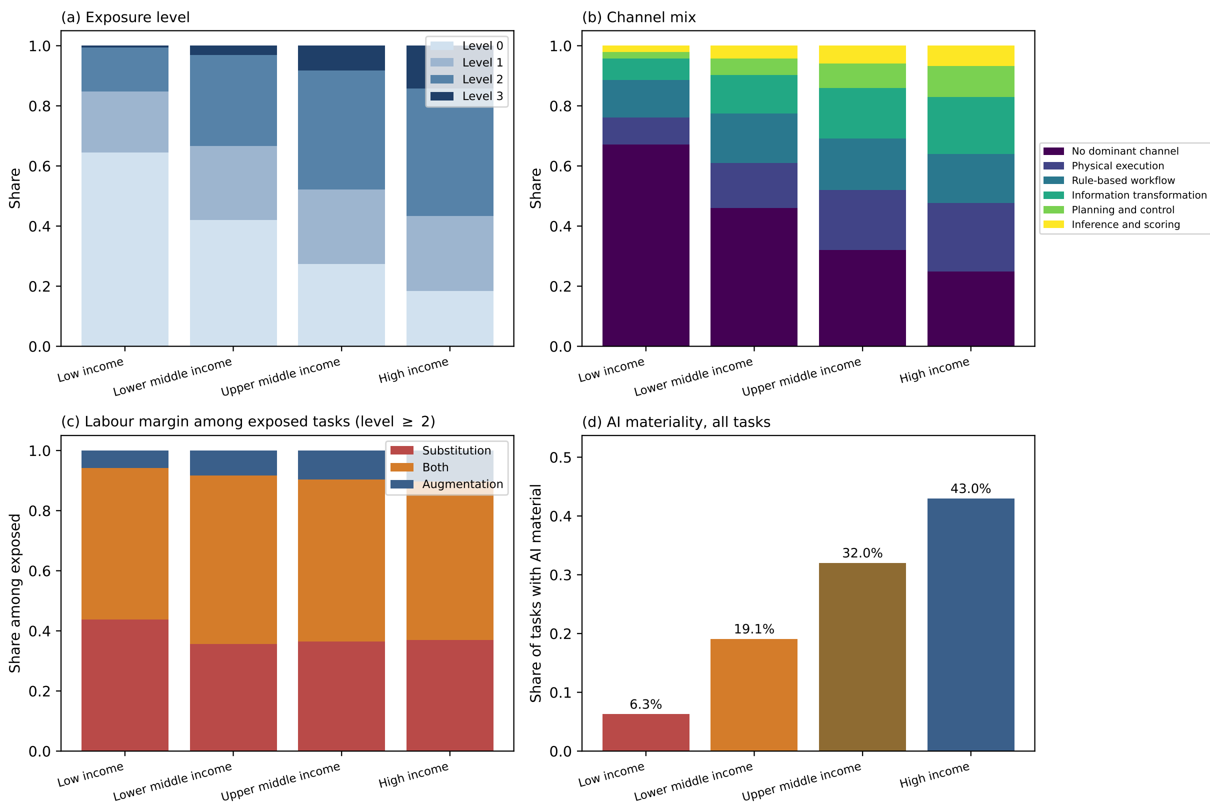
Notes: Panels report marginal distributions across the full 2,330,776-observation task-country dataset. Panel (a) reports exposure levels; panel (b) reports dominant technology channels; panel (c) reports labour margins among exposed tasks; panel (d) reports AI materiality. Exposure means level 2 or 3.

Table A.9: Illustrative rationale pairs from the country-conditioning divergence check.

Similarity pattern	Example task pair
Low Jaccard **Low cosine** *Substantively different* 24.7% of pairs	**Task:** Buy or sell stocks, bonds, commodity futures, or other securities on behalf of investment dealers. *Australia (high income):* Trading is highly automatable through algorithmic execution and AI-based predictive systems, which have largely displaced manual human order entry and routine portfolio management in the Australian financial sector. *Burundi (low income):* Burundi's current financial infrastructure lacks the digital market integration and high-volume electronic trading systems required for credible task-level automation of securities trading. Jaccard = 0.09, cosine = 0.36.
Low Jaccard **High cosine** *Same task, different reasoning* 74.9% of pairs	**Task:** Calculate costs for billings or commissions. *Saudi Arabia (high income):* Calculating costs, billings, and commissions is a deterministic, rule-based process involving structured financial data that is routinely and reliably automated by standard enterprise resource planning software. *Burundi (low income):* Calculating costs is inherently deterministic and standard business software in Burundi is already capable of automating these calculations, reducing the reliance on manual human arithmetic. Jaccard = 0.12, cosine = 0.60.
High Jaccard **Low cosine** *Rare edge* 0.0% of pairs	No observations fall in this region. A rationale cannot share a high share of content tokens with another rationale while being semantically unrelated, so the quadrant is empty by construction once both metrics are read together.
High Jaccard **High cosine** *Near-identical rationales* 0.4% of pairs	**Task:** Perform therapeutic wound care. *Greece (high income):* Therapeutic wound care requires patient-specific manual dexterity, clinical judgment, and direct physical interaction that cannot be reliably automated or meaningfully substituted by current technology. *Somalia (low income):* Therapeutic wound care requires high-level physical dexterity, clinical judgment, and patient interaction that cannot be reliably or safely automated in typical Somali clinical settings by 2026. Jaccard = 0.52, cosine = 0.79.

Notes: The table shows illustrative high-income/low-income rationale pairs selected to span low and high token overlap and low and high semantic similarity. Rationales are classifier justifications under country-conditioned prompts. Token overlap is measured with Jaccard similarity on content tokens; semantic similarity is measured with cosine similarity under OpenAI `text-embedding-3-large` embeddings.

Figure A.10: Label distributions by World Bank income group.



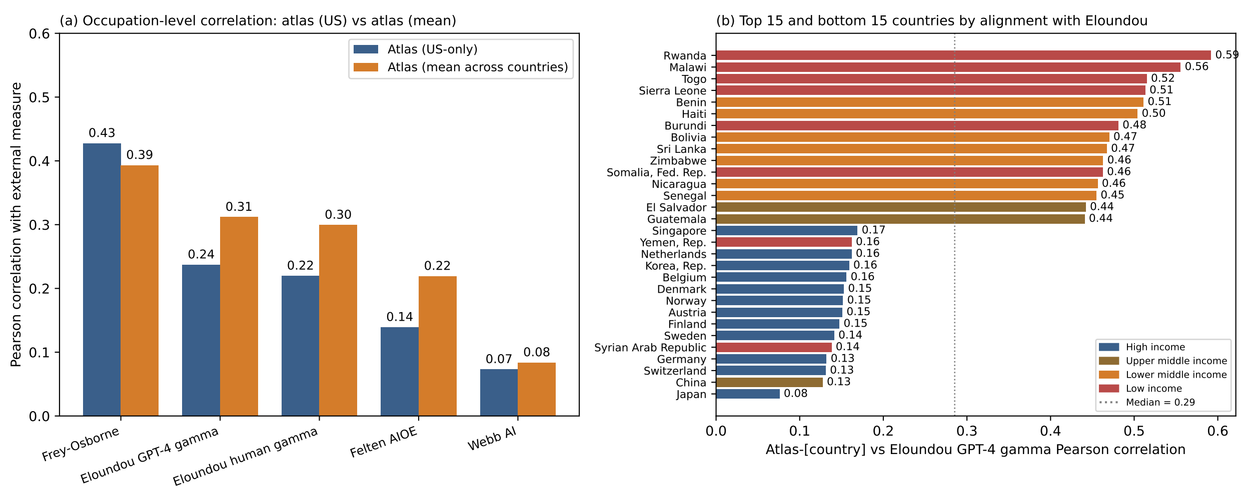
Notes: Panels report the same four label fields as Figure A.9, split by World Bank income group. Panel (a) reports exposure levels; panel (b) reports dominant technology channels; panel (c) reports labour margins among exposed tasks; panel (d) reports AI materiality. Exposure means level 2 or 3.

Supplementary Note A.6 What country conditioning adds

The construct-validity checks above show that the labels move with related external evidence. We now ask a different question: whether country conditioning changes the occupation exposure object, or whether similar patterns can be recovered from US-centric occupation scores plus country employment weights or a country-level scaling factor. Employment weighting explains part of the country ordering: an employment-weighted Felten AIOE score, built by weighting US-centric Felten occupation scores with country-specific ILOSTAT occupation shares, correlates with our country-conditioned AI-material share at Pearson 0.74 (Spearman 0.74) across 91 countries. But this still leaves about half of the variance outside the US-centric occupation-score-plus-employment-weight construction.

US-centric and context-free benchmarks. Figure A.11 compares our US-conditioned labels and our cross-country mean labels against external occupation scores. The US-conditioned version is closest to Frey–Osborne: the correlation is 0.427 using US labels and 0.393 using the cross-country mean. For Felten AIOE and both Eloundou gamma variants, the cross-country mean is closer: Eloundou GPT-4 gamma rises from 0.237 under US conditioning to 0.312 under the cross-country mean, and Felten AIOE rises from 0.139 to 0.219. Cross-country averaging behaves like a quasi-context-free summary, while country-conditioned labels carry institution-, industry-, and regulation-specific perturbations that a context-free score does not anticipate. Panel (b) shows the size of those perturbations: the correlation between each country's occupation-level exposure and Eloundou GPT-4 gamma ranges from 0.08 (Japan) to 0.59 (Rwanda), with a cross-country median of 0.29 and a USA value of 0.24. Low-income countries cluster at the top of the Eloundou-agreement ranking, while several large high-income economies cluster at the bottom.

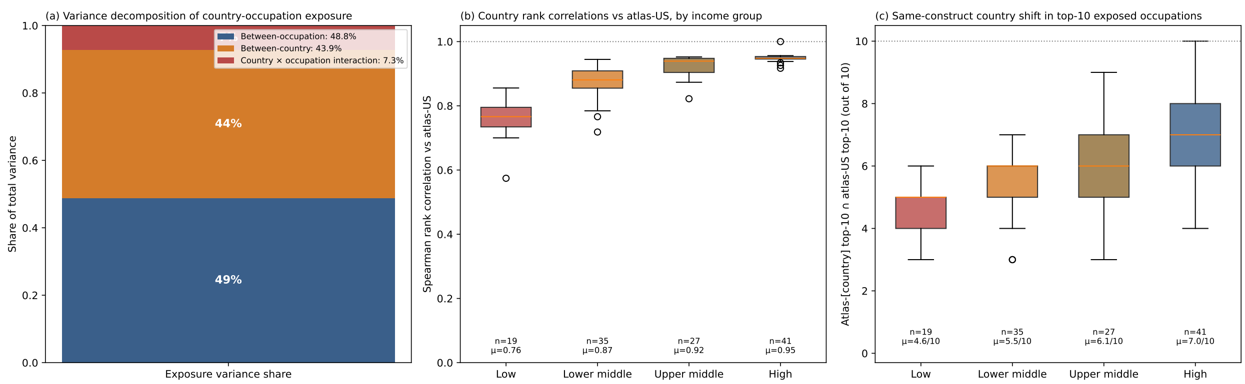
Figure A.11: Country conditioning relative to external occupation benchmarks.



Notes: Panel (a) compares occupation-level exposure correlations under US-conditioned labels and the unweighted cross-country mean. External measures are Frey–Osborne, Eloundou GPT-4 gamma and human gamma, Felten AIOE, and Webb AI exposure. Panel (b) reports each country's Pearson correlation with Eloundou GPT-4 gamma for the top and bottom 15 countries by alignment.

Country-conditioned occupation rankings. Figure A.12 quantifies the country-conditioned component directly. Panel (a) decomposes the 114,452 country–occupation observations into between-occupation, between-country, and country–occupation interaction variance. Occupations explain 48.8% of total variance and countries explain 43.9%, but the remaining 7.3% is country–occupation interaction: the component an additive construction of a US-centric occupation score and a country-level scaling factor would miss. For scale, an analogous ILOSTAT log-wage panel on the same country × occupation grid has only 0.9% country–occupation interaction variance. Panels (b) and (c) show where this component matters. The Spearman correlation between each country's occupation ranking and the US-conditioned ranking falls from 0.95 in high-income countries to 0.76 in low-income countries. Top-10 overlap falls in the same direction: high-income countries share 7.0 of the 10 occupations on average with the US-conditioned top-10 list, while low-income countries share only 4.6. Country conditioning therefore changes which occupations rise to the top of the exposure ranking, especially far from the US reference setting.

Figure A.12: Country conditioning beyond additive exposure scaling.



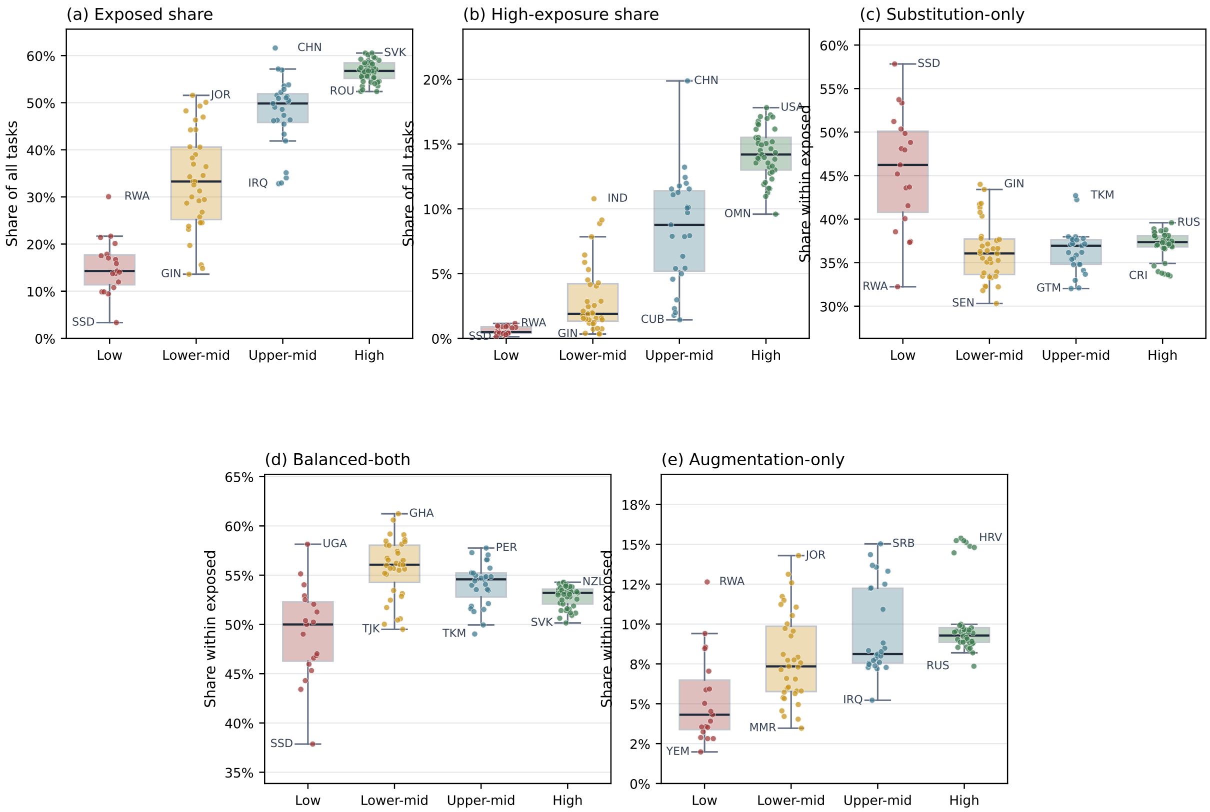
Notes: Panel (a) decomposes variance in the 124 country × 923 occupation exposure matrix into occupation, country, and country–occupation components. Panel (b) reports each country's Spearman rank correlation with the US-conditioned occupation ranking. Panel (c) reports overlap between each country's top-10 exposed occupations and the US-conditioned top-10 list.

The two comparisons give the same reading. Employment weights and cross-country averaging recover part of the country signal, while country conditioning adds more than a scalar adjustment to a US occupation score. It adds a country–occupation component that changes the exposure ranking, especially at the lower-income end of the development gradient. This finding aligns with cross-country heterogeneity in technology adoption and skill demand documented in Comin and Hobijn (2010), Alabdulkareem et al. (2018), Lewandowski et al. (2025), and Autor and Salomons (2018), and with evidence that the same occupation label can carry different task and skill content across economies and over time (Das and Hilgenstock, 2022; Deming, 2017).

Supplementary Note B Results Supplement

This note collects the figures and tables behind the Results: country exposure and within-income heterogeneity, margin and pathway companions, channel and AI-function companions, occupation and industry summaries, country-predictor robustness, and ILOSTAT employment-composition checks.

Figure B.1: Within-income-group variability across country-level summary metrics.



Notes: Each panel reports one country-level metric: exposed share (level ≥ 2, share of all tasks); high-exposure share (level $= 3$, share of all tasks); substitution-only share, balanced-both share, and augmentation-only share (each within exposed tasks). Box and whiskers report the within-income-group distribution across the 122-country sample; coloured points are countries; the highest and lowest country within each income group is labelled by ISO3 code. At fixed income tier, countries differ on all five metrics, with the widest within-tier spread in the lower-middle and upper-middle tiers.

Supplementary Note B.1 Country exposure and within-income heterogeneity

Table B.1: Country exposure by region and income tier.

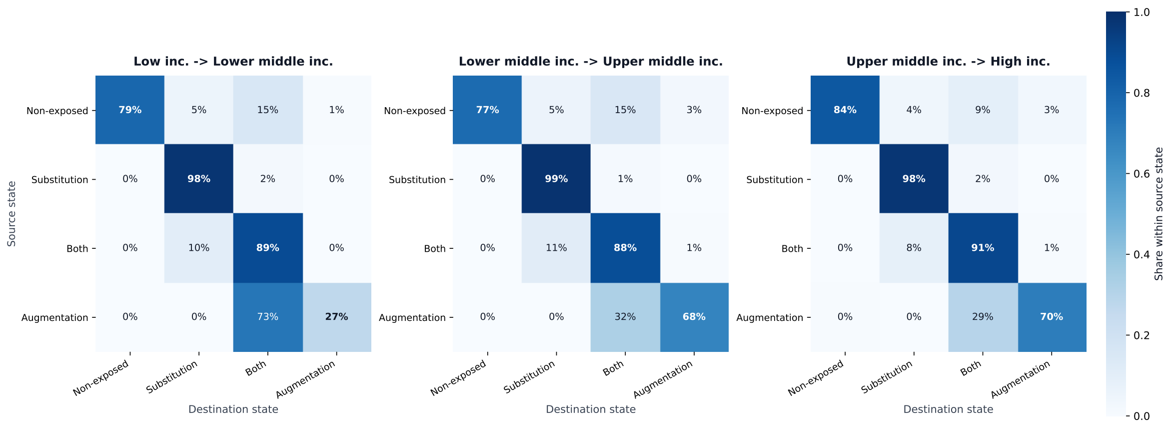
				Within exposed		
Income group	N	Exposed	High	Sub.	Both	Aug.
East Asia & Pacific						
High income	6	0.58	0.16	0.37	0.53	0.09
Upper middle income	4	0.53	0.13	0.38	0.54	0.08
Lower middle income	6	0.32	0.04	0.38	0.54	0.08
Europe & Central Asia						
High income	24	0.56	0.14	0.37	0.53	0.10
Upper middle income	7	0.51	0.09	0.37	0.52	0.11
Lower middle income	3	0.37	0.03	0.38	0.52	0.10
Latin America & Caribbean						
High income	3	0.57	0.12	0.34	0.51	0.14
Upper middle income	11	0.47	0.07	0.35	0.54	0.11
Lower middle income	4	0.32	0.02	0.34	0.55	0.11
Middle East & North Africa						
High income	6	0.56	0.13	0.36	0.52	0.11
Upper middle income	4	0.39	0.04	0.37	0.56	0.07
Lower middle income	4	0.48	0.06	0.33	0.55	0.11
Low income	2	0.16	0.01	0.52	0.45	0.02
North America						
High income	2	0.57	0.17	0.38	0.53	0.09
South Asia						
Lower middle income	5	0.38	0.05	0.35	0.57	0.08
Low income	1	0.14	0.01	0.50	0.47	0.04
Sub-Saharan Africa						
Upper middle income	1	0.51	0.12	0.38	0.55	0.07
Lower middle income	13	0.27	0.02	0.37	0.57	0.07
Low income	16	0.15	0.01	0.44	0.49	0.06

Notes: Exposed and high-exposure shares use all tasks as the denominator; substitution-only, balanced-both, and augmentation-only shares are computed within exposed tasks. Values are first computed by country and then averaged within region $\times$ income cells. Two countries without World Bank income classifications are omitted. Regional labels follow the paper's seven-region convention.

Supplementary Table B.1 reports the country-level exposure and margin summaries by region and income group. Read with Figure B.1, it shows that income carries much of the country exposure gradient, while meaningful within-tier dispersion remains.

Supplementary Note B.2 Margin and pathway companions

Figure B.2: Adjacent income-group transition matrices for all-task pathway states.



Notes: Each panel reports one adjacent income-group step, and rows sum to 100% within the source state. The sample is the same 18,797-task universe used in Figure 3b. Off-diagonal cells are tasks whose modal pathway state changes between adjacent income groups.

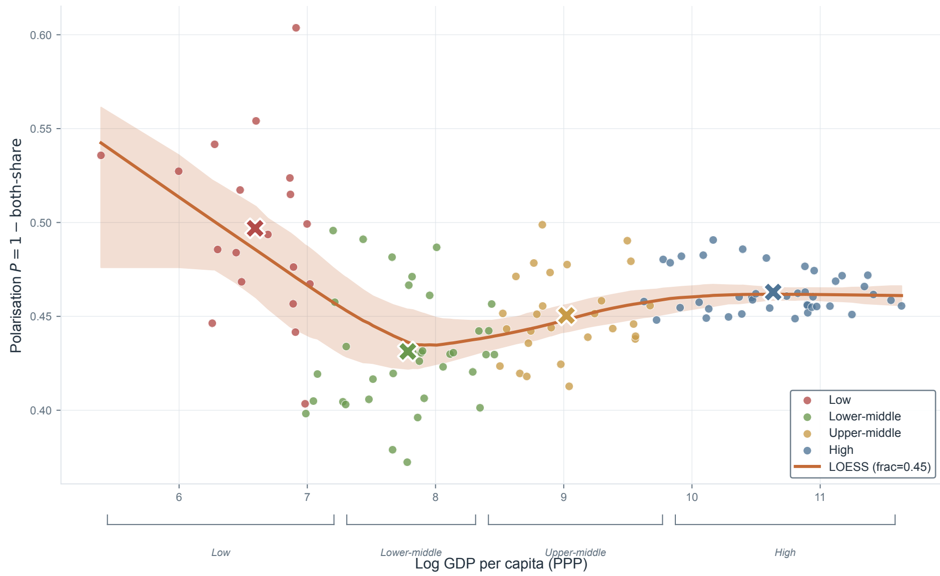
Supplementary Note B.2.1 Polarisation index as an alternative within-exposed summary

Within exposed work, each country has three mutually exclusive margin shares: substitution-only, balanced-both, and augmentation-only. We summarise the mass outside the balanced-both category with a polarisation index, $P_c = \text{sub}_c + \text{aug}_c = 1 - \text{bal}_c$. This adapts the idea of mass at the extremes from the labour-economics polarisation literature (Goos and Manning, 2007; Autor and Dorn, 2013). Here, the extremes are substitution and augmentation. The complementary tilt index $T_c = \text{sub}_c/(\text{sub}_c + \text{aug}_c)$ records which extreme dominates within the polarised mass. Because substitution dominates that mass in most countries, P is the more useful one-number summary.

Figure B.3 reports both the development gradient in polarisation P and its within-tier dispersion. The relationship is non-monotone: P is highest at the low-income end of the gradient (≈ 0.50), falls through lower-middle (≈ 0.43, the trough), and then rises gradually back to ≈ 0.46 across upper-middle and high-income countries. Low-income countries span roughly 0.40–0.60 on P (the widest spread); high-income countries span only 0.45–0.49 (the narrowest). Read together, the panels show that within-tier dispersion narrows as income rises, so margin polarisation is more stable among richer economies, where the balanced-both category absorbs a steadier share of exposed work.

Figure B.3: Polarisation P by development level and income tier.

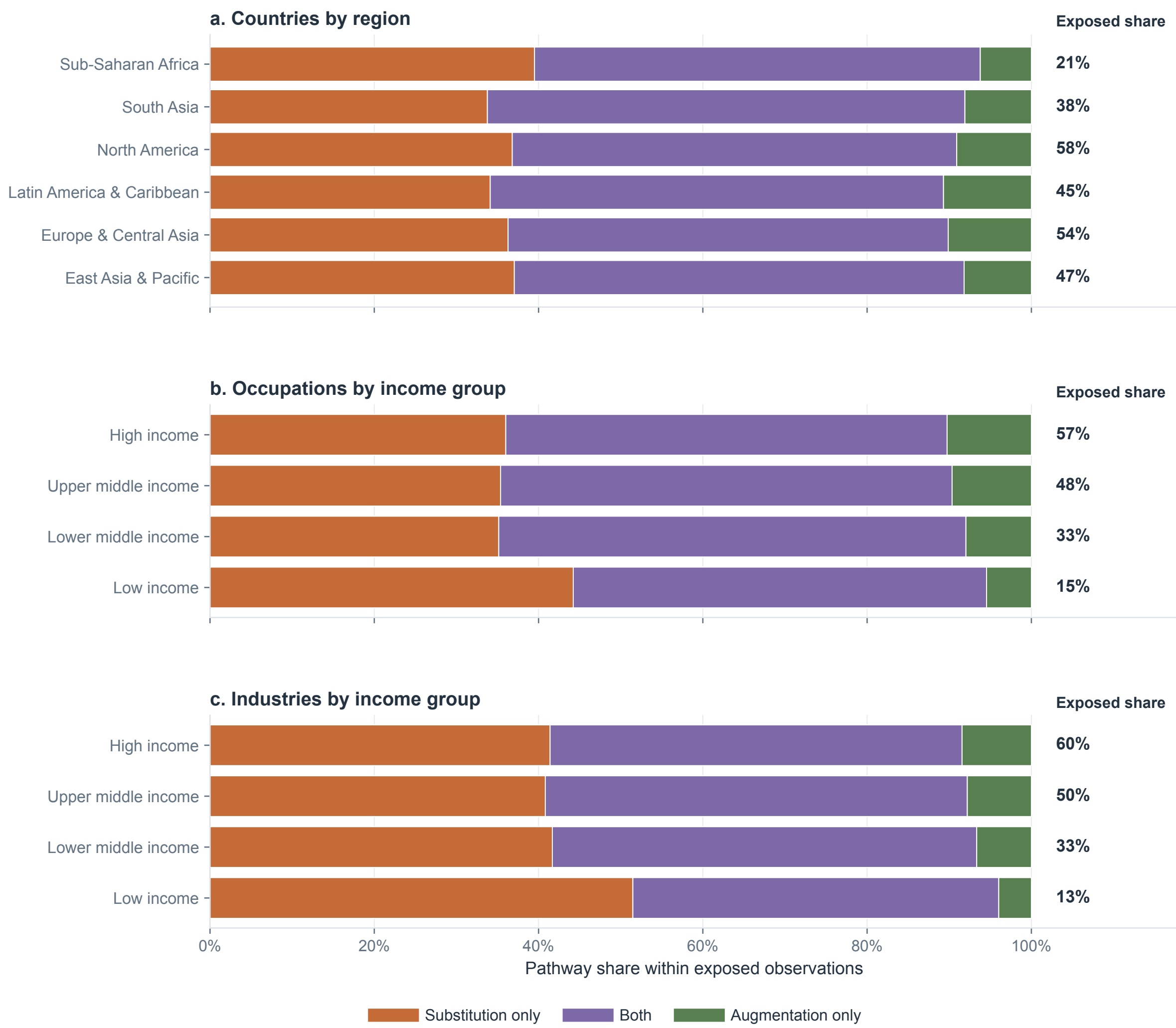
(a) Development gradient.



(b) Within-tier dispersion.

Notes: Polarisation is $P_c = \text{sub}_c + \text{aug}_c = 1 - \text{bal}_c$, where bal_c is the balanced-both share within exposed tasks. Panel (a) plots P against log GDP per capita with a LOESS smooth and 95% bootstrap interval from 200 country-level resamples. Panel (b) reports boxplots by World Bank income group, with country dots and labelled within-tier extremes.

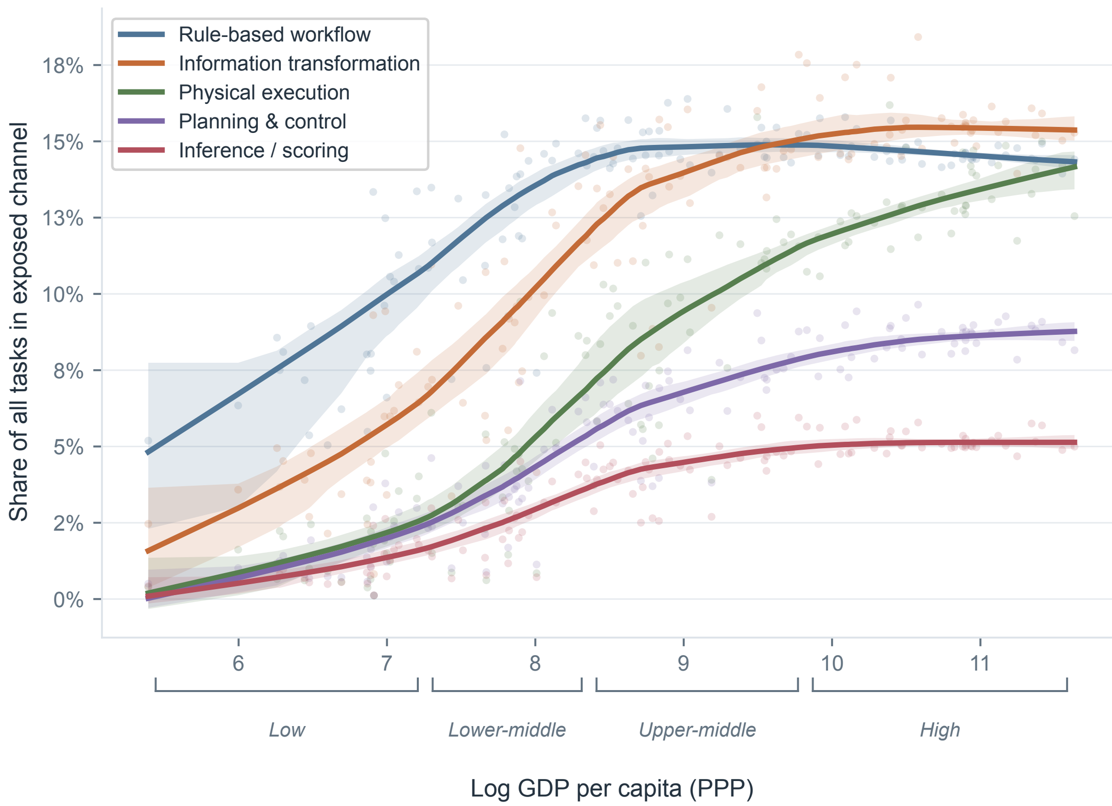
Figure B.4: Mutually exclusive pathway decomposition across country, occupation, and industry summaries.



Notes: Panel a reports regional country summaries, panel b reports the weighted occupation summaries by income group, and panel c reports the bottom-up industry summaries by income group. Bars partition exposed observations into substitution-only, balanced-both, and augmentation-only shares under the same mutually exclusive pathway definition used in the main text. Balanced-both exposure remains a large part of exposed work at all three reporting levels.

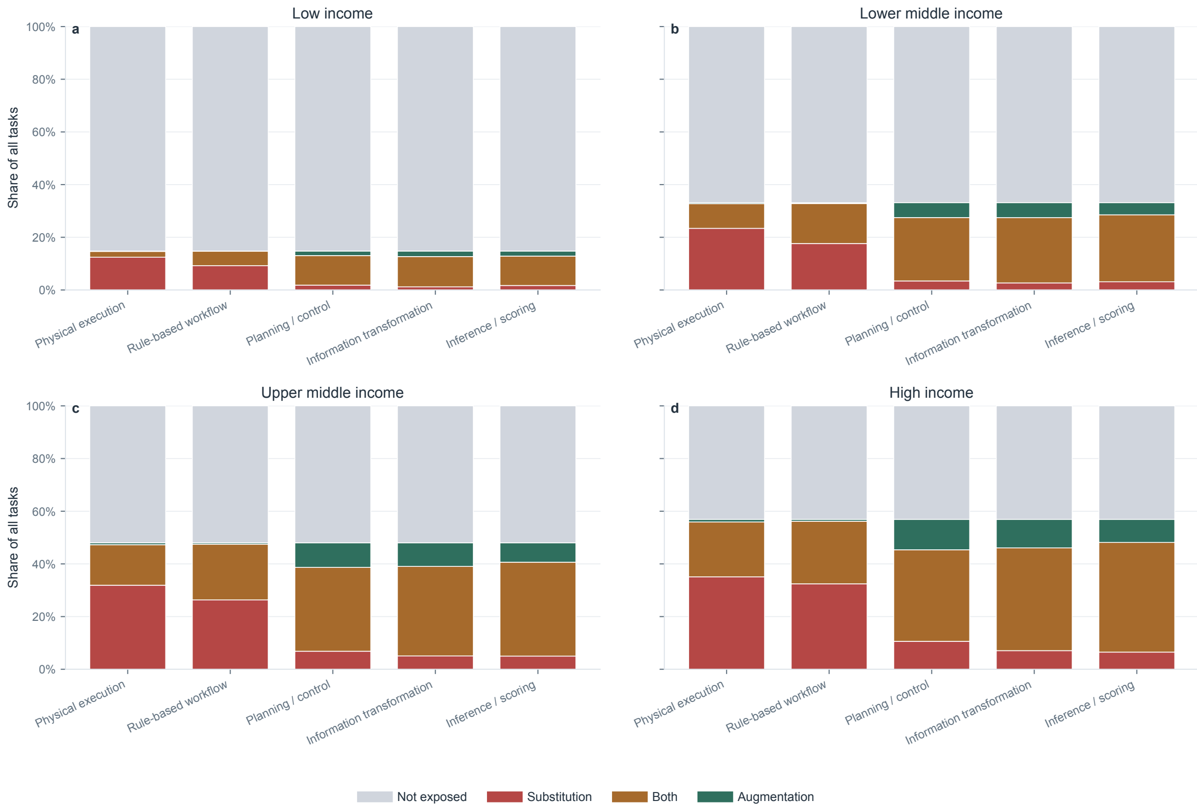
Supplementary Note B.3 Channel and AI-function companions

Figure B.5: Channel exposure as a share of all tasks across country development.



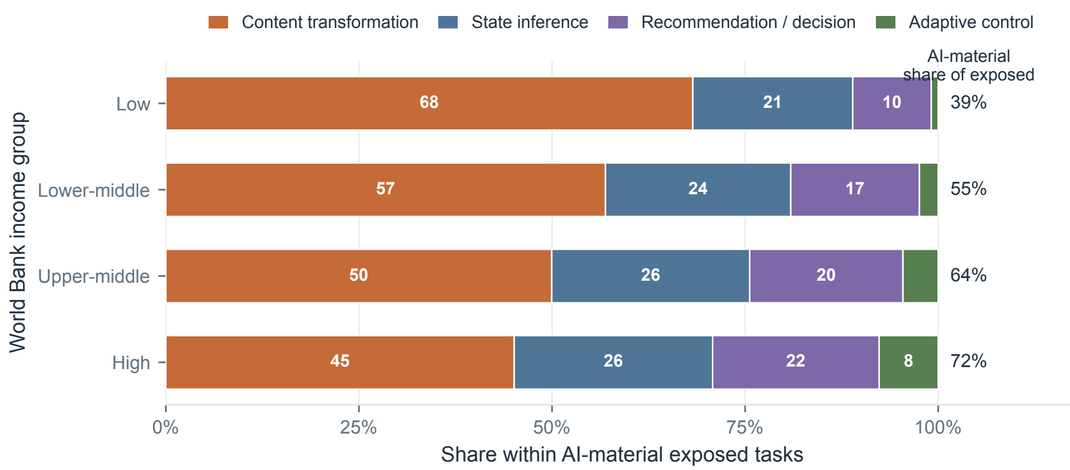
Notes: Lines report the country-level share of all tasks exposed through each dominant channel. Each line is a LOESS smooth against log GDP per capita over the 122-country sample with complete channel and GDP data. Shaded bands are 95% percentile bootstrap intervals from 200 country-level resamples; faded points are country observations.

Figure B.6: Channel composition of exposed work, broken out by World Bank income group.



Notes: Each panel reports one World Bank income group. Within each dominant channel, coloured segments show substitution-only, balanced-both, and augmentation-only shares; the grey segment is non-exposed task mass. Bars sum to 100% of tasks within each income group.

Figure B.7: AI-function mix within AI-material exposed tasks by income group.

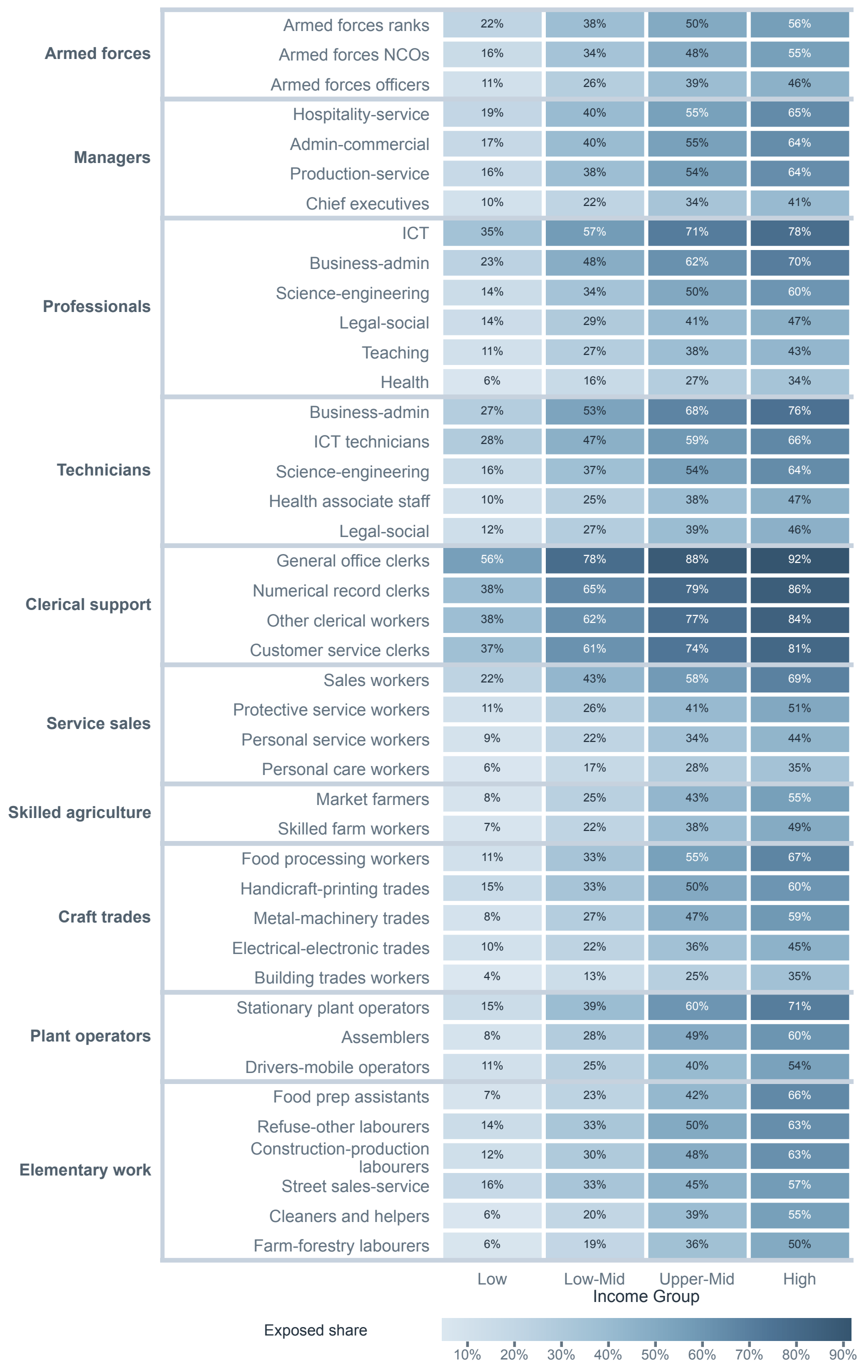


Notes: Rows are World Bank income groups. Stacked bars decompose AI-material exposed task-country observations into four dominant AI functions. Labels at right report the AI-material share among exposed observations in the same income group. The sample covers the 122 countries with classified income groups.

Supplementary Note B.4 Occupation and industry companions

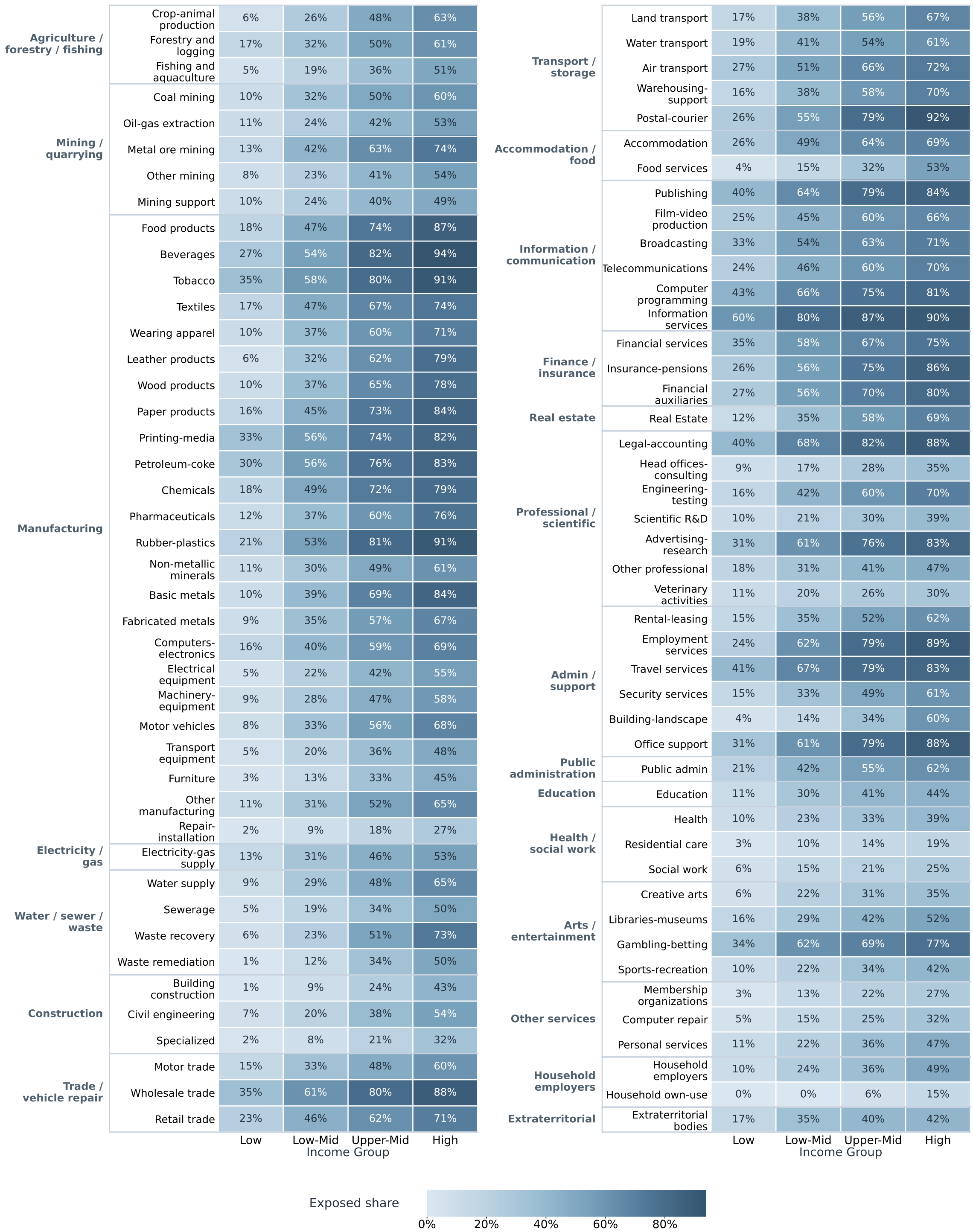
These figures aggregate the task-country labels into the ISCO and ISIC units referenced in Section 3.4.

Figure B.8: Weighted two-digit ISCO exposed shares by income group.



Notes: Columns are World Bank income groups. Rows are weighted two-digit ISCO occupation groups after SOC→ISCO linkage. Left tags mark one-digit ISCO major groups; outlined blocks show their constituent two-digit rows. Cells report exposed shares (the share of linked tasks at exposure levels 2 or 3 after linkage).

Figure B.9: Detailed two-digit ISIC task exposure by income group.



Notes: Rows are the 88 supported two-digit ISIC divisions, grouped by official section. Within each country, exposure is computed at the retained four-digit class level, averaged equally to two-digit divisions, and then averaged equally across countries within each income group. Cells report exposed shares at exposure levels 2 or 3.

Supplementary Note B.5 Top substitution and augmentation occupations and industries by income group

This subsection lists the occupation and industry rankings behind the main-text discussion of substitution-only and augmentation-only exposure. For each income group, the tables report the top entries selected by exposed share multiplied by the relevant labour-margin share among exposed tasks. The occupation table uses the weighted ISCO route; the industry table uses the bottom-up ISIC route used for the sector results. The purpose is descriptive: to show which occupations and sectors drive the substitution and augmentation patterns discussed in the main text.

Income group	Sub-only pocket					Aug-only pocket				
	Occupation	Exp.	Share	% AI	Channel	Occupation	Exp.	Share	% AI	Channel
Low	Billing and Posting Clerks	0.69	0.78	10	Rule workflow	Database Architects	0.45	0.13	70	Rule workflow
	Tellers	0.66	0.81	9	Rule workflow	Atmospheric and Space Scientists	0.31	0.13	81	Info trans.
	Payroll and Timekeeping Clerks	0.54	0.82	6	Rule workflow	Airline Pilots, Copilots, and...	0.23	0.18	31	Plan. control
Low/Mid	Tellers	0.90	0.77	24	Rule workflow	Air Traffic Controllers	0.37	0.28	62	Plan. control
	Billing and Posting Clerks	0.89	0.74	31	Rule workflow	Database Architects	0.67	0.15	80	Info trans.
	Brokerage Clerks	0.84	0.68	33	Rule workflow	Transportation Engineers	0.37	0.24	77	Info trans.
Upper/Mid	Tellers	0.97	0.80	32	Rule workflow	Air Traffic Controllers	0.51	0.32	75	Plan. control
	Billing and Posting Clerks	0.95	0.80	45	Rule workflow	Transportation Engineers	0.55	0.29	86	Info trans.
	Word Processors and Typists	0.87	0.80	51	Rule workflow	Electrical Engineers	0.56	0.26	87	Info trans.
High	Tellers	0.99	0.86	39	Rule workflow	Air Traffic Controllers	0.61	0.33	84	Plan. control
	Billing and Posting Clerks	0.95	0.86	65	Rule workflow	Transportation Engineers	0.65	0.30	93	Plan. control
	Word Processors and Typists	0.91	0.87	60	Rule workflow	Urban and Regional Planners	0.65	0.27	96	Info trans.

Table B.2: Top-three substitution-only and augmentation-only weighted ISCO occupations within each income group.

Note: Rows report the top three occupations on each side within each income group. Entries are selected separately using exposed share multiplied by the relevant pathway share among exposed tasks. Exp. is the linked occupation-level exposed share; Share is the sub-only or aug-only share within exposed tasks; % AI is the AI-material share among exposed tasks.

Income group	Sub. sector	Exp.	Share	% AI	Channel	Aug. sector	Exp.	Share	% AI	Channel
Low	Printing and media	0.25	0.52	35	Rule workflow	Apparel	0.21	0.05	29	Rule workflow
	Beverages	0.23	0.54	19	Rule workflow	Beverages	0.23	0.04	19	Rule workflow
	Apparel	0.21	0.52	29	Rule workflow	Other manufacturing	0.16	0.05	35	Rule workflow
Low/Mid	Printing and media	0.45	0.47	52	Rule workflow	Apparel	0.47	0.09	45	Rule workflow
	Chemicals	0.37	0.54	32	Rule workflow	Other manufacturing	0.35	0.09	44	Rule workflow
	Beverages	0.47	0.42	36	Rule workflow	Beverages	0.47	0.06	36	Rule workflow
Upper/Mid	Paper products	0.52	0.61	38	Physical exec.	Apparel	0.65	0.10	55	Rule workflow
	Chemicals	0.55	0.52	44	Physical exec.	Other manufacturing	0.51	0.10	55	Rule workflow
	Petroleum and coke	0.74	0.38	64	Physical exec.	Beverages	0.62	0.08	52	Rule workflow
High	Paper products	0.63	0.66	46	Physical exec.	Apparel	0.75	0.11	66	Info trans.
	Petroleum and coke	0.83	0.42	76	Physical exec.	Pharmaceuticals	0.73	0.10	66	Physical exec.
	Chemicals	0.66	0.52	56	Physical exec.	Beverages	0.71	0.10	68	Rule workflow

Table B.3: Top-three substitution-only and augmentation-only bottom-up ISIC sectors within each income group.

Note: Rows report the top three sectors on each side within each income group. Entries are selected separately using exposed share multiplied by the relevant pathway share among exposed tasks. % AI is the AI-material share among exposed tasks. The retained bottom-up ISIC graph is narrower than the occupation layer.

Supplementary Note B.6 Country-predictor robustness

These checks ask whether the country-predictor ranking changes with sample size, denominator, attribution rule, or correlated covariates. The appendix reports a residual-fit check in text and two figure companions: the earlier random-forest permutation-importance version and a linear Shapley R^2 companion for the 68-country predictor sample. Additional checks show that the wider-coverage 90-country specification preserves the same broad capability cluster, that adding the balanced-both margin to the within-exposed composition gives the same qualitative reading, and that winsorising the country covariates at the 1st/99th and 5th/95th percentiles leaves the leading capability cluster unchanged. For the augmentation-only outcome, the leading capital-intensity result is also unchanged in leave-one-country-out refits. These analyses describe country-level patterns and do not change the task labels. B.4 defines the country covariates used in the main and appendix screens.

As a residual-fit check, we regress raw country exposure and substitution-only share, and their GDP- or income-group residuals, on the same broad capability covariates. The structural covariate set explains the raw cross-country gradient well, with adjusted R^2 of about 0.93 for mean exposure and 0.92 for substitution-only share. Fit falls after conditioning on development stage: after removing income-group means, the same covariates explain about 0.11 of exposure residuals and 0.19 of substitution-only residuals. This supports the interpretation that the capability cluster captures much of the rich-poor gradient, while leaving substantial within-income country variation.

The wider-coverage 90-country specification addresses sample composition. The richer main-

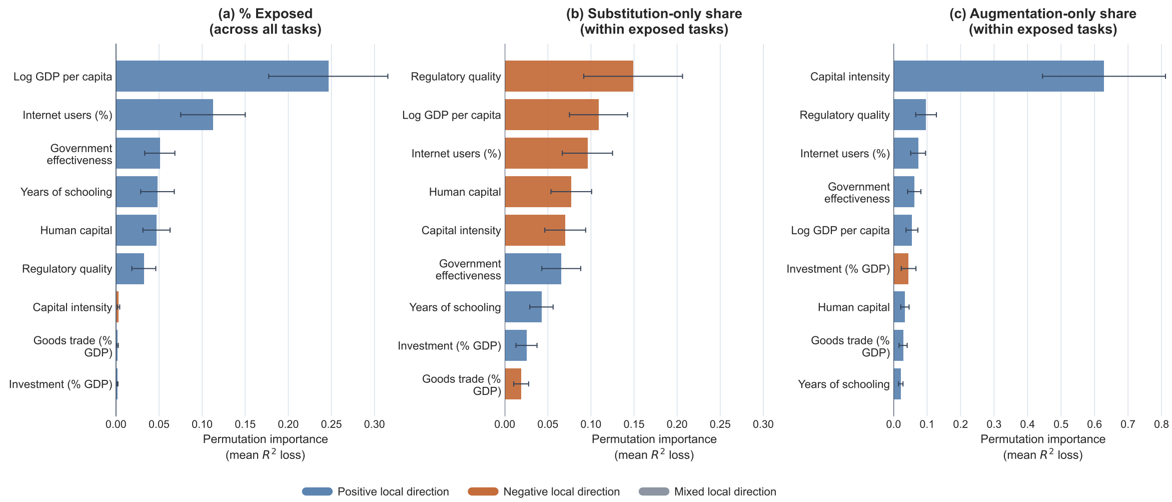
Table B.4: Country covariates used in the country-covariate screens.

Figure label	Construction	Source	Year rule
Log GDP per capita	Natural log of GDP divided by population, using the country-list GDP and population fields.	World Development Indicators (World Bank, 2024b)	2024
Human capital	PWT human-capital index, `hc`, based on schooling and returns to education.	Penn World Table 10.01 (Feenstra et al., 2015)	2019
Years of schooling	Average years of schooling among adults aged 15–64.	Barro–Lee educational-attainment data (Barro and Lee, 2013)	2015
Capital intensity	Natural log of real capital stock per worker, log(`rkna`/`emp`), where `emp` is persons engaged.	Penn World Table 10.01 (Feenstra et al., 2015)	2019
Investment (% GDP)	Gross fixed capital formation as a percentage of GDP.	World Development Indicators (World Bank, 2024b)	Latest non-missing, 2018–2024
Government effectiveness	Government effectiveness percentile rank.	Worldwide Governance Indicators (World Bank, 2024c)	2024
Regulatory quality	Regulatory quality percentile rank.	Worldwide Governance Indicators (World Bank, 2024c)	2024
Internet users (%)	Individuals using the Internet as a percentage of population.	World Development Indicators (World Bank, 2024b)	Latest non-missing, 2018–2024
Goods trade (% GDP)	Merchandise-trade value divided by GDP.	CEPII BACI HS22 2023 trade data (Gaulier and Zignago, 2010) and World Bank GDP (World Bank, 2024b)	2023 trade; latest GDP

Notes: The labels match the country-covariate figures. The main 68-country random-forest specification uses all rows in the table. The wider-coverage 90-country specification drops human capital, capital intensity, and investment to reduce complete-case sample loss. PWT variables are constructed national-accounts and education measures rather than survey-question variables. The PWT rows use 2019 values, the latest year in PWT 10.01 and a pre-COVID baseline. All logarithms are natural logs.

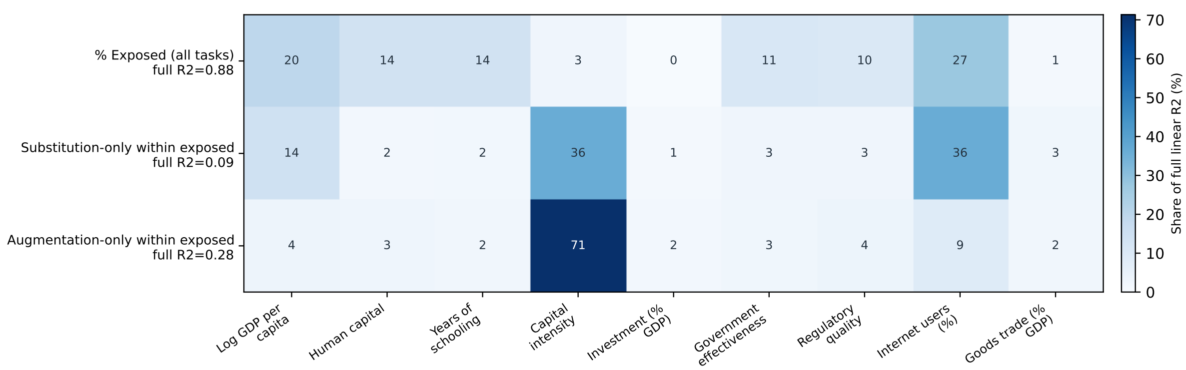
text specification retains more structural content, but its complete-case sample is smaller. Dropping the lowest-coverage structural variables expands country coverage and leaves GDP, digital connectivity, governance, and schooling as the dominant available cluster. A within-exposed version that adds the balanced-both margin gives the same qualitative reading: substitution-only and balanced-both shares track the development-capability cluster, while augmentation-only exposure is more closely associated with capital intensity.

Figure B.10: Permutation-importance companion for country predictors.



Notes: Bars report the earlier random-forest permutation-importance ranking in the same 68-country complete-case sample as Figure 6. Outcomes are exposed share among all tasks and substitution-only or augmentation-only shares within exposed tasks. Colours indicate the sign of the one-dimensional accumulated local effect. This companion is reported because unconditional permutation importance can be sensitive when predictors are correlated.

Figure B.11: Linear Shapley companion for the country-predictor specification.



Notes: Cells report exact dominance-analysis contributions from linear models on the same 68-country complete-case sample as the main country-predictor figure. Contributions average each covariate's incremental R^2 over all possible orderings. The outcomes are exposed share, substitution-only share within exposed tasks, and augmentation-only share within exposed tasks. This is a linear variance-decomposition companion, distinct from the TreeSHAP attribution used for the random-forest figures.

Supplementary Note B.7 Observed employment composition from ILOSTAT

This subsection asks how occupation exposure changes once observed worker shares replace task-linkage weights. Public ILOSTAT occupation data are available only at the ISCO-08 major-group level, so we collapse the weighted SOC→ISCO bridge to the same level before reweighting.

Supplementary Table B.5 reports the files used here, the country-year selection rules, and the resulting overlap with the task-country level dataset. The sex-by-occupation public series supports both total-employment reweighting and the female–male split used below.

Table B.5: ILOSTAT sources, sample rules, and overlap with the task-country dataset in the employment-composition appendix.

Source	Appendix object	Coverage rule	Usable countries	Paper overlap
SEX_OCU	Total occupation weights	Latest 2015–2025 year with total employment, at least 8 ISCO major groups, and positive counts; shares renormalized within country-year.	152	91
SEX_OCU	Female and male weights	Latest common 2015–2025 year with separate female and male counts, at least 8 ISCO major groups in each; shares renormalized within sex-country-year.	149	91

Note: 'Usable countries' reports how many countries satisfy the public ILOSTAT coverage rule for each source. 'Paper overlap' reports how many of those countries also appear in the task-country dataset used in the paper. The appendix figures use this overlap sample.

Figure B.12: Employment-weighted versus linkage-weighted occupation exposure by country.

Notes: Each point is a country with usable ILOSTAT total-employment occupation weights. The x-axis reports linkage-weighted exposed share from the task→SOC→ISCO bridge; the y-axis reports the same summary reweighted by observed ISCO-08 major-group employment shares. The 45-degree line marks equality.

Employment weighting preserves the broad occupation-layer gradient. High-income countries remain higher-exposure and low-income countries remain lower-exposure on average. But employment composition still moves some country summaries meaningfully because exposed major groups do not account for the same share of workers everywhere. Supplementary Table B.6 lists the countries where observed employment shares most raise or lower the occupation-layer exposed share relative to the linkage-weighted baseline.

Table B.6: Countries where observed employment composition most raises or lowers the occupation-layer exposed share.

Direction	Country	Income group	Linkage-weighted	Employment-weighted	Adjustment
Raised	Japan	High income	0.60	0.64	+0.04
	Korea, Rep.	High income	0.60	0.64	+0.04
	Hong Kong SAR, China	High income	0.56	0.59	+0.03
	Singapore	High income	0.61	0.63	+0.03
	Italy	High income	0.55	0.57	+0.02
	United Kingdom	High income	0.57	0.59	+0.02
Lowered	Rwanda	Low income	0.30	0.22	-0.09
	Benin	Lower middle income	0.25	0.16	-0.08
	Mozambique	Low income	0.19	0.11	-0.08
	Uganda	Low income	0.21	0.13	-0.08
	Cote d'Ivoire	Lower middle income	0.29	0.21	-0.08
	Zambia	Lower middle income	0.26	0.19	-0.08

Note: Positive adjustments mean that more exposed ISCO major groups account for a larger employment share than under the task-to-occupation linkage weights in that country.

The female–male split asks how exposure gaps look when the same task-country labels are reweighted by observed female and male employment shares. We report this as a data-linkage illustration rather than as a main gender result. Figure B.13 compares two public ILOSTAT views: two-digit ISCO-08 occupations and two-digit ISIC Rev. 4 industries.

For country c, sex s, cell j, and margin m, the employment-weighted contribution is

$$E_{csm} = \sum_{j} s_{cjs} x_{cjm},$$

where s_{cjs} is the employment share of sex s in occupation or industry cell j, and x_{cjm} is the corresponding country-conditioned exposure contribution for margin m. The plotted gap is $E_{cFm} - E_{cMm}$, so positive values indicate higher female employment-weighted exposure and negative values indicate higher male employment-weighted exposure. The occupation and industry readings differ. In two-digit occupations, substitution-only exposure is more female-weighted in the median country, while augmentation-only gaps are small. In two-digit industries, augmentation-only exposure is more male-weighted. This difference is informative: the same task-country labels can produce different gender gaps depending on whether employment is organised by occupations or industries.

To separate country-level accounting gaps from within-country sorting across cells, we also estimate cell-level fixed-effect regressions:

$$y_{cj} = \beta_m x_{cjm} + \alpha_c + \gamma_j + \varepsilon_{cj},$$

where y_{cj} is the female-minus-male employment share in occupation or industry cell j of country c, measured in percentage points; x_{cjm} is the corresponding exposure-margin share; α_c are country fixed effects; and γ_j are cell fixed effects. Each margin is estimated in a separate regression, and standard errors are clustered by country.

This exercise links the task-country labels to observed employment shares. It asks whether female and male workers are concentrated in different parts of the exposure distribution, and whether that answer changes when employment is read through occupations rather than industries. Occupation weighting reflects where women and men are employed across ISCO sub-major groups; industry weighting reflects where they are employed across ISIC divisions. The direction of the gap can therefore differ across the two views, especially for augmentation-only exposure.

The fixed-effect table asks a narrower sorting question. Conditional on country averages and on the average gender composition of each occupation or industry cell, are more exposed cells

Figure B.13: Female–male exposure gaps by labour margin, occupation and industry weighting.

Notes: The x-axis reports the female-minus-male employment-weighted contribution in percentage points. Positive values indicate higher female employment-weighted exposure; negative values indicate higher male employment-weighted exposure. Points show country-group medians and horizontal lines show interquartile ranges. Occupation uses sex-specific ILOSTAT employment shares by two-digit ISCO-08 sub-major group. Industry uses sex-specific ILOSTAT employment shares by two-digit ISIC Rev. 4 division. Exposure scores are held fixed within each occupation or industry cell; differences come from observed female and male employment composition.

Table B.7: Cell-level gender sorting and automation margins.

	(1) Occupation sub.	(2) Occupation aug.	(3) Industry sub.	(4) Industry aug.
Exposure margin (10 pp)	-0.351***	0.273	-0.219***	0.008
	(0.119)	(0.393)	(0.041)	(0.074)
Observations	3,084	3,084	4,837	4,837
Countries	89	89	73	73
Cells	42	42	88	88
Country fixed effects	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes	Yes

Notes: The outcome is the female-minus-male employment share in a country–cell, measured in percentage points. Each column is a separate regression. Exposure margins are scaled so coefficients correspond to a 10 percentage point increase. Occupation cells are two-digit ISCO-08 sub-major groups; industry cells are two-digit ISIC Rev. 4 divisions. Standard errors, clustered by country, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

more female- or male-intensive? Substitution-intensive cells employ relatively more men in both occupation and industry specifications. The augmentation coefficients are smaller and statistically imprecise, so the evidence is strongest for gender sorting along the substitution margin.

This reading is consistent with the broader gender-exposure literature in a more limited sense: headline AI exposure can be gender-neutral or female-tilted in occupation-based data (Pizzinelli et al., 2023; Cazzaniga et al., 2025; Gmyrek et al., 2023, 2025), while older automation and robotics evidence often emphasizes male-intensive routine and production work (Acemoglu and Restrepo, 2018; Brussevich et al., 2018, 2019). Our contribution here is not a new gender theory; it is to show that the task-country labels can be reweighted through observed employment shares and decomposed by labour margin.