Automation without Borders: Generative AI and Cross-Border Service Outsourcing

Neha V. Betai Maggie X. Chen*

February 3, 2026

Abstract

The outsourcing of routine cognitive work to developing countries has been a resilient component of globalization. This paper investigates how generative artificial intelligence (AI) reshapes cross-border service outsourcing by tracing firm and worker responses to the diffusion of AI in a global market for digital services. Using granular transaction data linking clients, workers, and skills from 2022–2024, together with a novel capability-based measure of AI automation exposure spanning 2,700 skill categories, we document a sharp contraction in cross-border outsourcing in highly exposed tasks. Outsourcing volume falls by roughly 35 percent, with declines concentrated in developing-country destinations that previously specialized in codifiable work. These aggregate effects mask substantial reorganization. On the demand side, firms consolidate remaining outsourcing into fewer, larger, and more complex contracts. On the supply side, exposed workers complete fewer contracts but earn more per engagement, as a subset reallocates toward AI- and interaction-intensive tasks. Adjustment is uneven, leading to a 26-percentage-point increase in within-skill concentration of activity. A task-based trade framework rationalizes these patterns by showing how generative AI compresses cross-border wage and productivity arbitrages in routine tasks, shifting comparative advantage and redistributing gains from trade in digital services.

JEL codes: F1, O3

Keywords: Generative AI, international outsourcing, services, reskilling, concentration

*George Washington University. Neha Betai: nehabetai96@gwu.edu; Maggie Chen (Corresponding author): xchen@gwu.edu. We thank seminar and conference participants at Yale's 2025 Firms, Trade, and Development Conference, Paris School of Economics "AI, Digitization and Trade Patterns" Workshop, Inter-American Development Bank, World Bank, International Finance Corporation, HKUST-Fudan-NUS Conference on International Economics, and Lingnan-Nottingham Workshop on International Trade for valuable comments and suggestions.

1 Introduction

The rise of generative artificial intelligence (AI) marks a potential turning point in the global organization of work. Over recent decades, advances in communication technology enabled firms in developed countries to offshore routine cognitive tasks to workers in developing countries, where wages were lower but skills were sufficient. This "trade in tasks" reshaped labor markets on both sides of the transaction, creating export opportunities for workers in developing economies while enabling cost savings and specialization for firms in developed economies (Grossman and Rossi-Hansberg, 2008; Baldwin, 2019). Generative AI threatens to upend this equilibrium. By automating precisely the routine, codifiable tasks that underpinned developing countries' comparative advantage, the technology may compress the cross-border wage arbitrage that sustained global digital outsourcing, altering not only how work is performed, but where it is produced.

This paper provides the first evidence on how generative AI reshapes cross-border service outsourcing. We ask how generative AI-driven automation alters the volume, composition, and geography of outsourced work, and how firms and workers on both sides of the market adjust in response. Understanding these dynamics is essential for assessing the distributional consequences of AI-driven automation: unlike earlier waves of technological change that primarily displaced workers in developed countries, generative AI directly targets tasks that had become globally tradable and that supported export-led growth in parts of the developing world. As a result, the adjustment costs of automation may be reallocated across countries as well as across workers.

We address these questions using transaction data from a global digital labor platform specialized in cross-border outsourcing, where buyers, predominantly in developed countries, post tasks and workers in developing countries bid and complete contracts, providing a direct window into international trade in digital services. The setting offers three key advantages for studying how AI reshapes international outsourcing and its labor market consequences. First, we observe both sides of the market adjusting simultaneously—firm reorganization and worker adaptation—rather than inferring one from the other. Second, we track adjustment in real time, observing the evolution of outsourcing behavior from nearly a year before ChatGPT's release through two full years of diffusion. Third, the platform spans more than 2,700 narrowly defined skill categories with varying exposure to AI automation, enabling us to exploit rich cross-skill variation in a difference-in-differences design while tracing reallocation across the full task space.

We combine these microdata with a novel, skill-level measure of AI automation exposure. Rather than relying on occupation-level proxies or realized adoption patterns, we directly

elicit generative AI's self-assessed capability to perform each of the platform's 2,700 skill categories, mirroring how firms themselves evaluate automation potential by testing frontier models on representative tasks. The resulting index exhibits strong validity against internal and external benchmarks, and captures within-occupation heterogeneity that coarser measures average away. This capability-based approach identifies what AI can technologically perform, isolating the scope for automation from the organizational frictions that shape equilibrium adoption.

We develop a conceptual task-trade framework to guide the empirical analysis and clarify the underlying mechanisms. Buyers allocate tasks among domestic production, offshore outsourcing, and AI automation; workers with heterogeneous ability choose which skills to supply and whether to invest in reskilling. Generative AI reduces the cost of automating routine cognitive tasks, compressing the wage-productivity arbitrage that previously sustained offshore outsourcing. The framework generates four sets of predictions: aggregate contraction of international outsourcing concentrated in developing-country destinations specialized in routine, codifiable tasks, buyer task consolidation into complex and AI-complementary projects, directed worker reskilling with differential success by ability, and rising within-skill concentration as threshold effects amplify inequality among those who remain.

We empirically investigate the above hypotheses and document large effects that propagate through both sides of the market, from aggregate contraction to firm reorganization to worker reskilling, and trace their distributional consequences. First, international outsourcing to developing countries declines sharply in skills exposed to generative AI automation. At mean exposure, outsourcing volume falls by 35 percent, with the number of active buyers and workers declining in parallel. Crucially, displacement is concentrated in developing-country destinations. Outsourcing to developed countries, where pre-AI comparative advantage lay in complex, less codifiable tasks, shows no significant decline, consistent with the framework's prediction that AI compresses the wage arbitrage sustaining developing-country specialization in routine digital services.

Second, buyers do not simply reduce international outsourcing; they fundamentally reorganize it. Exposed buyers consolidate remaining work into fewer, larger, and more complex contracts: total outsourcing value rises by 46 percent among continuing buyers at mean exposure, while the number of contracts falls by 12 percent, implying a 65 percent increase in average contract value. This consolidation is accompanied by longer project durations, more detailed task specifications, and wider budget spreads, consistent with a shift toward work that is harder to automate and more difficult to price. Exposed buyers also increase their demand for projects explicitly referencing AI integration and deployment. These patterns suggest that generative AI does not simply displace outsourcing but transforms its compo-

sition: buyers withdraw routine, codifiable tasks while retaining and expanding work that requires human judgment, coordination, and oversight.

Third, workers face displacement but a subset actively adapts. Exposed workers complete 37 percent fewer contracts but earn 41 percent more per engagement, indicating a compositional shift toward higher-value work rather than uniform contraction. Beneath this average response lies substantial skill reallocation. High-exposure workers reduce the breadth of skills they supply while simultaneously adopting new competencies, with transitions directed toward AI-complementary and interaction-intensive tasks that retain value in an automated environment. Critically, this adaptation is highly uneven across the ability distribution. Experienced and high-rated workers are significantly more likely to add new skills, including AI-specific skills, and significantly less likely to shed existing competencies; less established workers more often exit or contract their portfolios without successfully reorienting. The returns to adaptation are large: workers who reskill into AI-complementary domains see earnings gains exceeding 60 percent, while those who fail to reorient experience stagnant or declining earnings.

Fourth, these heterogeneous responses generate sharp distributional consequences. The concentration of market opportunities rises by 26 percentage points in high-exposure skills, as top performers capture an increasing share of a shrinking market. This increase occurs primarily within skill categories rather than through reallocation across them; high-exposure skills were already more concentrated before AI, so their contraction slightly reduces aggregate concentration. What drives the increase is rising inequality among workers who remain active in the same skill domain. Worker turnover actually reduces measured concentration, as top performers are more likely to reskill and enter new, AI-complementary domains. Workers in exposed skills thus face a dual challenge: aggregate demand declines, and within the reduced market, opportunities concentrate among those best positioned to adapt.

These findings show that generative AI reshapes globalization not by uniformly reducing demand for digital services, but by reorganizing where and how such services are produced. By compressing the wage and productivity differentials that sustained offshoring of routine tasks, AI withdraws a core margin of trade in tasks while simultaneously expanding demand for complex, human-intensive work. Adjustment therefore occurs through reorganization and selection rather than across-the-board displacement. The resulting rise in within-skill concentration reflects higher effective capability thresholds rather than market frictions, indicating that inequality can increase even in flexible, competitive labor markets. From a development perspective, the results caution that digital services—often viewed as a durable pathway to export-led growth—may be especially vulnerable to automation, shifting adjustment costs toward workers in developing countries. This pattern suggests that development

strategies centered on digital outsourcing may need to place greater emphasis on adaptable skill formation, task upgrading, and institutions that facilitate worker transition as generative AI reshapes the tradability of services.

Related literature. This paper contributes to the rapidly growing literature on generative AI and work while shifting the lens from domestic employment and within-job productivity to the reorganization of internationally traded digital services. A pioneering body of work quantifies the potential exposure of occupations to large language models and other AI technologies (Eloundou et al., 2024; Felten et al., 2019; Lou et al., 2023), measures AI usage through interaction or task-level data (Handa et al., 2025), and documents productivity effects through experimental and quasi-experimental designs (Noy and Zhang, 2023; Brynjolfsson et al., 2025; Dell'Acqua et al., 2023). We complement these important contributions by moving beyond occupational exposure and within-task productivity to study how generative AI reshapes the organization of production across borders, documenting changes in realized outsourcing, task composition, and worker skill allocation in a global market.

Building on earlier work that established online labor markets as an important setting for studying labor market interactions and contracting (e.g., Horton, 2010; Agrawal et al., 2016), a second strand offers novel evidence on how generative AI affects contractors using platform data. Hui et al. (2024) document short-run displacement effects on Upwork, while Teutloff et al. (2025) show declining demand for AI-substitutable skills alongside growth in AI-complementary roles. Our analysis differs along three dimensions. First, we examine both demand and supply sides of the market—buyer reorganization and worker adjustment—within a single setting. Second, our platform specializes in cross-border outsourcing, enabling us to examine how generative AI reshapes international trade in services and the labor markets of traditional outsourcing destinations, a distinct margin from studies of domestic platforms. Third, we move beyond displacement to unpack the underlying adjustments from buyer consolidation to directed reskilling and rising concentration, linking demand-side restructuring directly to worker-level outcomes and distributional consequences.

Our analysis also relates to the broader literature on the economic implications of artificial intelligence and automation. Earlier work emphasizes that AI differs from traditional automation by affecting a wide range of tasks without fully automating entire occupations (Brynjolfsson et al., 2018). Evidence from job postings documents rising demand for AI-related skills among exposed firms (Alekseeva et al., 2021; Acemoglu et al., 2022), while firm- and occupation-level studies find limited aggregate employment effects despite substantial reorganization within firms and jobs (Acemoglu et al., 2022). We add to this literature by showing that muted aggregate employment effects can coexist with substantial reorganiza-

tion of production across borders, as generative AI reshapes the allocation of tasks, earnings, and opportunities through international outsourcing.

The paper also contributes to the literature on trade in tasks and the organization of global production (Grossman and Rossi-Hansberg, 2008, 2012; Antràs and Chor, 2013; Antràs, 2015) by providing direct, transaction-level evidence on how a general-purpose technology reshapes the geography and structure of internationally traded services. While earlier work emphasizes how declines in trade, coordination, or communication costs expand offshoring and task fragmentation, our results show that advances in automation can operate in the opposite direction, compressing cross-border wage and productivity arbitrages and withdrawing routine tasks from international outsourcing. In this sense, generative AI alters comparative advantage through endogenous reallocation of tasks and production stages, consistent with models of evolving specialization and global value chain adjustment (Baldwin, 2019). In contrast to prior studies of automation in developing countries, which focus primarily on robotics and manufacturing (Faber, 2020; Artuc et al., 2023; DeStefano and Timmis, 2024), we study automation in digital services, a sector often viewed as a pathway to export-led growth in the global South, and trace how adjustment costs are distributed across workers through structured reskilling and rising concentration.

The remainder of the paper proceeds as follows. Section 2 describes the cross-border online labor platform and the linked buyer-worker-skill transaction data. Section 3 describes the construction of our skill-level measure of exposure to generative AI. Section 4 presents a task-based conceptual framework formalized in Appendix A that motivates the empirical analysis and clarifies the mechanisms linking AI automation to outsourcing and worker adjustment. Section 5 documents aggregate effects of generative AI on cross-border outsourcing. Section 6 examines how exposed buyers reorganize outsourcing, focusing on changes in contract structure, project complexity, and task composition. Section 7 analyzes worker-side responses, including activity, reskilling, earnings dynamics, and the resulting distributional consequences. Section 8 concludes.

2 Empirical Setting and Data: Cross-Border Online Labor Market

Online labor marketplaces provide a rare micro-level view of trade in services by matching firms to workers worldwide for digitally deliverable tasks, a segment of trade that has ex-

panded rapidly since the mid-2010s.[1] Our empirical analysis uses data from a leading global online labor marketplace (OLM) that specializes in cross-border trade in digitally deliverable services.[2] Unlike many other platforms, which facilitate primarily domestic contracting or offer limited transaction observability, this platform focuses on international outsourcing and fully mediates project-based contracts between firms and workers located in different countries. The data cover a broad range of activities organized into more than 2,700 skill categories, covering software development, web design, data analysis, marketing, writing and translation, business services, and more. Crucially, the platform records the universe of posted projects, bids, matches, and realized transactions, allowing us to observe detailed information on task content, contract terms, prices paid, work outcomes, and the geographic locations of both buyers and sellers.

These features also distinguish our data from job-posting datasets such as Burning Glass, which capture intended vacancies rather than realized employment relationships. The institutional structure of this platform allows us to observe both sides of the market and equilibrium outcomes, including selection into contracts, pricing, and the cross-country allocation of tasks. These data therefore provide an unusual lens on high-frequency adjustments in international outsourcing across extensive, intensive, and compositional margins that are largely unobservable in conventional trade and labor datasets.

2.1 Key Features of the Platform

The defining characteristic of the platform in our study is its strong cross-border orientation. In our data, 91 percent of contracts involve a client and a worker located in different countries, with demand originating primarily in high-income economies, led by the United States, and supply concentrated in developing countries, particularly South Asia. This geographic structure reflects the platform's role as infrastructure for international trade in services rather than domestic gig work. It allows us to directly observe how AI-driven technological change affects the cross-country allocation of tasks, a margin that is central to debates on globalization and development but difficult to study using conventional data sources.

Compared to other OLMs, the platform also serves a broader and more heterogeneous set of clients, including multinational corporations, government agencies, small and medium-sized enterprises (SMEs), and individual entrepreneurs. This diversity distinguishes our set-

[1] Two broad types of platforms operate in the gig economy: (i) location-based platforms (e.g., ride-sharing and local services) that require physical delivery, and (ii) remote work platforms (e.g., digitally deliverable professional services) where output can be produced and delivered online. Because generative AI primarily affects digitally produced content and code, we focus on the latter.

[2] We do not name the platform due to the terms of our data access agreement. Appendix B provides additional institutional details about the setting.

ting from enterprise-focused outsourcing platforms that disproportionately serve large firms with established procurement processes. The inclusion of SMEs is substantively important, as smaller firms often lack in-house capacity for specialized digital tasks and may be particularly responsive to changes in the relative cost, quality, and availability of external labor. As a result, the data capture AI-driven reorganization across the firm size distribution rather than only among large enterprises.

The platform organizes work into more than 2,700 skill categories. This fine-grained task structure enables measurement of AI exposure at the skill level and allows us to trace reallocation across narrowly defined activities, rather than across occupations or industries. Projects range from routine digital tasks, such as data entry or basic translation, to highly specialized services, including software architecture and machine learning implementation, generating substantial variation in task complexity and automation exposure.

Further, all contracts on the platform are mediated through a decentralized matching process that ensures comprehensive observability of outcomes. Clients post project listings specifying task requirements, budgets, and timelines, and workers submit proposals describing qualifications, experience, and compensation terms. Clients select among applicants based on observable characteristics such as price, skills, and prior performance. Once a contract is formed, the platform facilitates communication, milestone tracking, and payment, and records completion and performance feedback. This process generates a consistent and complete project-level record of outsourcing relationships and outcomes.

2.2 Data: Workers, Clients, and Transactions

We collect worker-, client-, and contract-level transaction data for all 2,700 skill categories on the platform from January 2022 through December 2024. A central feature of online labor marketplaces is that the population of registered users vastly exceeds the population of active participants: only a small fraction of workers ever complete a contract, and an even smaller fraction do so regularly. Economically relevant outcomes including outsourcing volumes, task composition, and worker earnings are generated almost entirely by the subset of workers who are actively matched to projects. Our analysis focuses on this active segment, capturing the margin of digital services trade that responds to demand shocks rather than the universe of registered but inactive accounts.[3]

To implement this focus, we construct a fixed cohort of workers with demonstrated engagement prior to the diffusion of generative AI. For each granular skill, we sample up to

[3]Throughout the paper, we interpret outcomes as changes in observed outsourcing activity within the platform marketplace. Our analysis is not designed to capture firms' in-house labor adjustments or changes in total labor demand outside the platform.

500 workers who completed at least one contract in that skill before November 2022, based on the platform's internal ranking system which orders the search results by the number of completed contracts. Across the 2,700 skills, more than 1,500 are highly specialized.[4] Over 75 percent of the platform's granular skill categories have fewer than 500 active workers; for these skills, typically narrowly specialized categories, our sample includes the universe of participants. For the remaining 25 percent categories, our sample captures the visible portion of the market: workers outside the top within a skill rarely appear in buyer searches. Further, workers frequently list multiple skills; the median worker in our sample appears in 8 skill categories.

This procedure identifies around 93,000 unique individuals. Given that our analysis requires observing workers both before and after November 2022, the final sample consists of approximately 17,000 workers who meet this criterion. The sampling strategy is both economically meaningful and empirically conservative. In the ten months prior to November 2022, these continuously active workers in our sample captured approximately 95 percent of contract volume, consistent with the well-documented concentration of gig economy platforms, where a small core of participants generates most transactions.[5] Including lower-ranked workers would add noise without materially affecting measured outsourcing patterns.[6]

For each worker in the cohort, we collect the complete history of completed projects visible on their platform profile between January 2022 and December 2024. The resulting dataset contains about 300,000 projects. Consistent with the platform's cross-border orientation documented above, the vast majority involve parties in different countries.[7]

The final dataset consists of linked client-worker-skill transaction microdata for cross-border services trade. Worker profiles provide location, experience, education, and certifications (when available), prior job history, and client ratings. Client records include country, hiring history, industry classifications, and ratings. Project-level data contain titles and descriptions, associated skill categories, contract status, budget ranges, hours worked, observed payments, and multidimensional performance ratings covering quality, communication, timeliness, and professionalism.[8] The structure of the data allows us to trace high-frequency

[4]These include advanced engineering and science skills such as astrophysics and bioinformatics, new emerging technologies such as LoRA Training, and legacy software such as COBOL.

[5]See Horton (2010) and Agrawal et al. (2016) on concentration in online labor markets.

[6]Fixing the cohort ensures that changes in employment, earnings, or task composition are not mechanically driven by shifts in platform traffic. To the extent that AI deters entry by workers who would have joined absent the technology, our estimates may understate aggregate displacement effects.

[7]Some projects display limited payment information (e.g., contracts marked as private or sealed). We retain these projects for outcomes that do not require payment information. Differences in sample sizes across specifications are documented.

[8]Appendix B provides additional details on the data collection and variable definitions.

adjustments in outsourcing behavior, task allocation, and worker outcomes in response to generative AI. Because we observe realized contracts rather than intended demand, we can detect both immediate and persistent changes in client and labor behavior that are largely invisible in aggregate labor market and trade statistics.

3 Measuring Skill-Specific Exposure to Generative AI Automation

We construct a skill-specific measure of AI exposure that captures heterogeneity in generative AI capabilities across the platform's 2,700 skill categories. The measure is designed to reflect how firms evaluate automation potential in practice: by directly testing frontier models on representative tasks.

Definition and measurement. We define AI exposure as the extent to which generative AI can perform the tasks associated with a given skill, thereby reducing the need for human labor. This capability-based concept abstracts from firm-specific adoption frictions and organizational constraints, instead measuring the technological scope for automation as perceived by market participants at the time decisions are made.

For each skill that does not require physical presence, we elicit assessments of generative AI capabilities using structured prompts. Specifically, we provide the model with tasks drawn from the platform's skill descriptions and request a quantitative feasibility rating:

> "I am going to list the tasks for which a contractor could be hired under the category of [skill]. On a scale from 0 to 1, how well can generative AI perform these tasks relative to a human, with minimal human input?"

The baseline index uses GPT-3.5, the first large language model to achieve widespread commercial adoption during our sample period. This choice aligns the measure with the capabilities available when firms began adjusting outsourcing decisions. To reduce sampling variation, we repeat each prompt 100 times in independent API sessions (temperature = 1, no conversation history) and average across iterations.[9]

Validation. Table 1 reports skills with the highest and lowest AI exposure. The ranking exhibits strong face validity. High-exposure skills center on text-based, rule-governed tasks such as programming, translation, proofreading, data entry for which LLMs can perform

[9] Appendix C documents the full prompt text, API parameters, and internal and external validations.

core activities with minimal human input. Low-exposure skills involve obsolete technologies and highly specialized domains requiring tacit knowledge or context-specific judgment.

The ranking also reveals substantial within-field heterogeneity. For example, while general programming ranks among the most exposed skills, legacy technologies (e.g., COBOL) appear near the bottom, reflecting limited training data and narrower model applicability. This pattern suggests that exposure is driven not by technical sophistication per se, but by the extent to which tasks can be specified through codifiable instructions and evaluated using standardized criteria.

We benchmark our skill-level measure against occupation-based (and when possible, task-based) AI exposure indices from Eloundou et al. (2024) and the Anthropic AI Exposure Index. To implement these measures, we map each platform skill category either to O*NET tasks using TF-IDF cosine similarity between skill names and task descriptions and then assigning similarity-weighted average exposure scores or to occupations using a standardized occupation crosswalk. Appendix C.3 reports these comparisons and Appendix D.3 re-estimates our main specifications using each alternative measure.

Relation to existing measures. Our approach differs from prior methodologies along three dimensions. First, it operates at the level of skills and captures the margin along which reallocation occurs in digital outsourcing markets where contracts specify narrowly defined tasks and technological change is more likely to reshape the composition of work within fields than to uniformly expand or contract entire occupations. Second, it measures technological capability rather than realized usage, separating what AI can do from how firms choose to deploy it.[10] Third, it reflects how economic agents probe AI in practice through direct model interaction, making the measure a behavioral proxy for firms' beliefs about automation potential.

A potential concern is that a language model evaluates its own capabilities. However, this mirrors firm behavior: managers assess automatability by testing models on representative tasks. To the extent that self-assessment errors are uncorrelated with true exposure, measurement error attenuates estimated effects toward zero. Moreover, the 2,700 skills span domains over which no group of human evaluators possesses uniform expertise. Human-based assessment would inevitably embed discipline-specific priors that are not commensurate across fields. LLM elicitation thus provides a practical alternative to infeasible comprehensive human assessment.

[10] Measures based on application logs or user prompts are informative about equilibrium deployment but may conflate capability with adoption decisions. Our approach isolates the automation frontier.

Complementary skill measures. Beyond exposure to automation, we construct three additional measures to capture reallocation across skill types. First, we construct an alternative *assistance* exposure measure based on AI-assisted task execution rather than full automation. The prompt asks how well generative AI can assist an individual in performing the tasks associated with a given skill, capturing settings in which human input remains necessary. Appendix D.3 reports the analysis using this alternative measure.

Second, we identify *AI skills*: categories explicitly referencing AI integration, prompt engineering, model deployment, or workflow automation. Activity in these skills captures upgrading and specialization that may accompany displacement.

Third, we measure *interactive intensity*: the share of projects within each skill whose descriptions reference communication, client interaction, creativity, or coordination. This captures dimensions of work relying on interpersonal judgment that remain less amenable to automation even when the underlying task is digitally deliverable.

These measures allow us to distinguish three margins of adjustment: displacement from automatable tasks, expansion into AI-complementary work, and reallocation toward interaction-intensive services. In the empirical analysis, we use these measures to assess whether observed changes in outsourcing and task composition reflect substitution away from automatable work, upgrading toward AI-complementary technical skills, and reallocation toward interactive tasks that remain human-intensive.

4 Conceptual Framework

This section develops a task-based framework linking generative AI to changes in international service outsourcing. The framework builds on canonical models of task-based production and offshoring (Grossman and Rossi-Hansberg, 2008, 2012) and on theories of task-replacing technological change (Acemoglu and Restrepo, 2018, 2020). Appendix A provides a formal model; here we summarize the key mechanisms and derive testable predictions.

We characterize tasks along two related dimensions: their exposure to AI automation and their complexity. Exposure reflects the extent to which generative AI can perform a task autonomously, while complexity captures the degree to which a task requires coordination, customization, or judgment. In digital services markets, these dimensions are negatively related. Tasks with high AI exposure, such as routine coding, translation, standardized content production, data processing, are modular and codifiable: they can be fully specified ex ante and verified against objective criteria. Tasks with high complexity, such as system architecture, strategic consulting, and creative direction, require ongoing interaction, tacit

knowledge, and judgment that resists codification.[11]

These task characteristics interact with cross-country differences in wages and productivity to shape pre-AI outsourcing patterns. Developing countries' lower wages confer a comparative advantage in routine, codifiable tasks where labor cost is the primary determinant of sourcing decisions. More complex tasks, in contrast, favor workers in developed countries, whose productivity advantages in coordination, communication, and judgment can offset higher wages. As a result, prior to the diffusion of generative AI, tasks with high automation exposure are disproportionately outsourced to developing countries, while developed-country contractors specialize in lower-exposure, higher-complexity work, consistent with task-based theories of offshoring and comparative advantage (Grossman and Rossi-Hansberg, 2008, 2012).

The public release of large language models in late 2022 constituted a discrete improvement in the quality and availability of AI tools capable of performing text-based, code-based, and design-adjacent tasks. For buyers, this improvement reduces the effective cost of performing high-exposure tasks algorithmically.[12] Tasks that were previously outsourced because of low developing-country wages become candidates for in-house automation. Because high-exposure tasks are concentrated in developing-country outsourcing, this substitution disproportionately displaces contractors in developing countries. Outsourcing to developed countries, concentrated on more complex tasks that remain below the automation threshold, is comparatively insulated.

As automatable tasks exit the outsourced set, the composition of remaining outsourcing changes. The residual work becomes increasingly complex, interdependent, and difficult to specify ex ante. Such tasks benefit from relational contracting: they require coordination across subtasks, adaptation over time, and accumulated client knowledge. Buyers therefore respond by consolidating outsourcing into fewer, larger contracts with contractors capable of handling broader project scope and sustained interaction.

On the supply side, these changes alter the relative value of skills. Skills tied to automatable tasks decline in value, while skills that facilitate coordination, integration, and client interaction, including AI-complementary and interactive skills, become more valuable as buyers reorganize toward complex projects. Workers reallocate in response, with transitions concentrated among technologically proximate skills and directed toward higher-value activities. High-ability workers, who face lower costs of reskilling and adaptation, are better

[11] This negative relationship between codifiability and complexity is a standard feature of task-based models (Autor et al., 2003; Acemoglu and Restrepo, 2018) and of the relational contracting literature (Antràs, 2015).

[12] We examine in Appendix A alternative channels through which generative AI may affect task trade, including productivity augmentation and task expansion and discuss their respective theoretical and empirical implications.

positioned to respond to these shifts, consistent with models of task-specific human capital and differential adjustment to technological change (e.g., Acemoglu, 1998).

These forces jointly increase concentration in outsourcing outcomes. Selection operates as contractors specialized in displaced tasks exit. Threshold effects arise as consolidation raises the capability requirements for remaining contracts. Differential adaptation further amplifies concentration, as high-ability workers are more likely to transition successfully into AI-complementary roles.

The framework yields four predictions that discipline our empirical analysis. Appendix A formalizes these mechanisms and derives each prediction.

Prediction 1: Task displacement. As AI capabilities improve, outsourcing volume declines more sharply in skills with higher AI exposure. This follows directly from the substitution channel: tasks that AI can perform autonomously exit the outsourced set.

Prediction 2: Destination asymmetry. The decline in outsourcing is concentrated among contractors in developing countries. This follows from pre-AI sorting: because developing countries specialized in high-exposure tasks, they bear the brunt of AI-induced displacement.

Prediction 3: Contract consolidation. Continuing buyers shift toward fewer, larger, more complex contracts. We therefore expect increases in average contract value, contract duration, and interactive task content.

Prediction 4: Worker adjustment and concentration. Workers, especially high-ability ones, reallocate toward AI- and interaction-intensive skills. Concentration increases as select workers capture disproportionate shares of residual contracts.

We test these predictions in the following sections by exploiting variation in exposure to generative AI across skills, businesses, and workers, the timing of major model releases, and the rich transaction-level data from the online labor marketplace, which allow us to observe realized contracts, client task reallocation, and worker skill adjustment at high frequency.

5 Aggregate Effects on International Outsourcing

We begin by examining the first two predictions of the framework in Section 4 and Appendix A at the aggregate level: whether advances in generative AI lead to differential changes in international outsourcing across skills with varying degrees of automation exposure. Focusing on aggregate outcomes provides a starting point: it allows us to assess whether AI reshapes overall outsourcing volumes and cross-country patterns in ways consistent with

task-level substitution, before investigating how these changes are mediated through buyer behavior and worker adjustment.

Empirical specification. Our empirical strategy is a difference-in-differences design that exploits two sources of variation: (i) the timing of the arrival and diffusion of generative AI, and (ii) cross-skill heterogeneity in the technological scope for automation, measured by our time-invariant exposure index $Exposure_i$ constructed in Section 3. Intuitively, we compare changes in outsourcing before and after the onset of the generative-AI era for skills that are more versus less exposed to automation.

We estimate the following baseline specification at the skill–week level:

$$y_{it} = \gamma_i + \lambda_t + \beta\big(Exposure_i \times PostGPT_t\big) + \varepsilon_{it}, \tag{1}$$

where i indexes skill subcategories and t indexes weeks. The outcome y_{it} denotes measures of outsourcing activity in skill i at time t, including the total value and number of completed contracts. The indicator $PostGPT_t$ equals one for weeks after the onset of widespread access to generative AI (our baseline cutoff is the public release of ChatGPT), and zero beforehand. Skill fixed effects γ_i absorb time-invariant differences across skills (including baseline demand, typical contract size, and market thickness). Week fixed effects λ_t flexibly control for common shocks to the platform such as seasonality, aggregate demand fluctuations, and macroeconomic conditions. Standard errors are clustered at the skill level to account for serial correlation in outcomes within skills over time.

The coefficient β captures whether outsourcing changes differentially after the introduction of generative AI in skills with greater automation exposure. Because $Exposure_i$ is time-invariant, this is a continuous-treatment difference-in-differences design: identification comes from differential changes over time across skills as a function of their ex ante exposure to AI automation. Appendix D.1 reports estimates using exposure bins rather than the continuous measure.

The key identifying assumption is that, absent generative AI, outsourcing in high- and low-exposure skills would have followed parallel trends conditional on skill and time fixed effects. We assess this assumption using an event-study specification that allows the exposure gradient to vary flexibly by month:

$$y_{it} = \gamma_i + \lambda_t + \sum_{\tau \neq -1} \beta_\tau \big(Exposure_i \times \mathbb{1}\{t = \tau\}\big) + \varepsilon_{it}, \tag{2}$$

where τ indexes months relative to November 2022. Pre-period coefficients $\{\beta_\tau\}_{\tau<0}$ provide a direct test for differential pre-trends; post-period coefficients trace the dynamic evolution

of exposure-dependent effects.

Our design does not require that firms or workers immediately adopt generative AI. We interpret the release of ChatGPT and subsequent model upgrades (GPT-4, GPT-4o) as discrete capability shocks that expand the technological frontier for task automation. The difference-in-differences design identifies how outsourcing responds to changes in what AI can perform, rather than to realized adoption. This distinction matters: even if take-up is gradual or heterogeneous, reorganization of production must ultimately reflect changes in technological capability. Our exposure-based design isolates this channel by comparing skills that differ in their susceptibility to AI-driven automation.

We estimate our baseline specification on a panel of skill-week observations from January 2022 to December 2024. Our main analysis focuses on international outsourcing to developing countries, which accounts for 93 percent of cross-border transactions on the platform.[13]

Aggregate effects. Table 2 reports the effects of generative AI on international outsourcing at the skill–week level. Consistent with the first prediction of the conceptual framework in Section 4, we find large and statistically significant declines in outsourcing concentrated in skills with greater exposure to AI automation. Column (1) reports estimates for log total contract value: the coefficient on $Exposure_i \times PostGPT_t$ is -0.820, implying that a skill at mean exposure (0.69) experienced a decline in outsourcing value of approximately 43 percent relative to the pre-ChatGPT period. Column (2) shows a corresponding decline in contract volume of about 35 percent at mean exposure. Together, these estimates indicate a substantial contraction in both the scale and frequency of outsourced work in tasks where AI can most readily substitute for human labor.

Columns (3) and (4) decompose this contraction into participation margins.[14] Both the number of buyers posting contracts and the number of workers completing contracts decline significantly with exposure. At mean exposure, market participation falls by roughly 34–35 percent on both sides. The parallel response of buyers and workers is consistent with a demand-driven mechanism in which firms withdraw highly automatable tasks from the market, reducing opportunities for contractors in exposed skill categories.[15]

The conceptual framework further predicts that displacement can be less pronounced in developed countries, where comparative advantage prior to AI diffusion was strongest in complex, less codifiable tasks. Figure 4 evaluates this prediction by comparing responses

[13] We examine outsourcing to developed countries separately later in this section.

[14] Sections 6 and 7 explore the intensive margins among continuing buyers and workers.

[15] We interpret these results as capturing reorganization of production within the platform. Declines in outsourcing may reflect in-house automation, re-internalization of previously outsourced tasks, or shifts to alternative platforms. We cannot fully disentangle these channels with aggregate data; Sections 6 and 7 provide finer evidence on mechanisms by tracking buyer and worker behavior directly.

across destination countries. The outsourcing response in developed countries is closer to zero and statistically insignificant, in sharp contrast to the large negative effects observed for developing-country destinations. This asymmetry aligns with the framework's logic: developing-country contractors specialize in high-exposure tasks that are automated first, while developed-country outsourcing remains concentrated in more complex tasks that remain below the automation threshold.[16]

Figures 1 and 2 examine the dynamics underlying these aggregate effects. Across all outcomes—total contract value, number of projects, active buyers, and active workers—event-study coefficients are flat and statistically indistinguishable from zero prior to the diffusion of generative AI, providing no evidence of differential pre-trends between high- and low-exposure skills. The exposure gradient emerges gradually following the release of GPT-3.5 and becomes markedly more negative after the release of GPT-4 and GPT-4o. The common timing of these shifts across value, volume, and participation margins indicates that the contraction in outsourcing coincides with discrete improvements in AI capability rather than with pre-existing differences across skills.

Appendix D reports various specification and robustness checks. First, the results are stable to alternative functional forms, including the inverse hyperbolic sine transformation to accommodate zeros and the use of exposure bins instead of a continuous measure. Second, the estimates are qualitatively similar using the Eloundou et al. (2024) α and β measures, the Anthropic AI Exposure Index, and our alternative measure based on AI-assisted production rather than full automation. While these external indices provide useful validation of the aggregate results, they cannot substitute for our skill-level measure in the buyer and worker analyses that follow: quantifying adjustments by market participants requires exposure defined at the level of platform-specific skills that map directly to observed contracts, task bundles, and worker portfolios.[17]

[16]The imprecision of the developed-country estimate may also reflect the small volume of outsourcing to these destinations, just 7 percent of cross-border contract value in the pre-period. This compositional difference is itself a manifestation of comparative advantage: prior to AI diffusion, low-wage countries captured the bulk of routine, codifiable tasks that are most amenable to outsourcing. The limited developed-country sample thus provides less statistical power but reinforces the economic interpretation.

[17]We also show in Appendix D.5 that the magnitude of the response varies systematically across destination countries. The contraction is largest in Bangladesh and India, countries that, prior to generative AI, specialized most heavily in routine, codifiable digital services. In contrast, effects are smaller and less precisely estimated for countries such as Egypt, where baseline outsourcing volumes are lower and skill composition is more heterogeneous. This pattern suggests that rather than uniformly affecting all destinations, generative AI disproportionately erodes demand for tasks that formed the core of comparative advantage in certain developing-country labor markets.

6 Buyer and Task Reorganization

The aggregate contraction in outsourcing documented above can mask substantial reorganization in what buyers continue to contract and how they structure outsourced work. Improvements in generative AI may lead firms not only to substitute automation for routine tasks, but also to consolidate remaining tasks into fewer projects that are broader in scope and require greater coordination across skills. We therefore examine whether buyers more exposed to generative AI reorganize outsourcing along these margins.

Our analysis is conducted at the buyer-week level. For each outcome, we estimate

$$y_{bt} = \gamma_b + \lambda_{ct} + \beta\big(\text{Exposure}_b \times \text{PostGPT}_t\big) + \varepsilon_{bt}, \tag{3}$$

where γ_b are buyer fixed effects capturing buyers' time-invariant characteristics and λ_{ct} are source-country-week fixed effects absorbing source country-specific labor demand and supply shocks. Buyer Exposure_b is constructed using the buyer's pre-GPT outsourcing portfolio: specifically, we take the maximum automation exposure of all skills outsourced by buyer b prior to November 30, 2022. This measure captures the buyer's ex ante vulnerability to AI-driven substitution and is fixed before treatment.[18] We study buyer reorganization along two margins: (i) project structure and spending, and (ii) task composition and complexity.

6.1 Outsourcing Structure and Budgets

We first test whether exposure to generative AI is associated with contraction and consolidation in buyer contracting behavior, as predicted by the framework in Section 4 and Appendix A. Table 3 presents estimates of equation (3) for buyers who remain active on the platform after the introduction of ChatGPT.[19] The results reveal systematic reorganization among high-exposure buyers. Column (1) shows that total outsourcing value increases by 46 percent for buyers with a mean automation exposure of 0.52. However, column (2) indicates that this increase does not reflect more frequent contracting: the number of contracts declines by 12 percent. Combining these estimates, column (3) confirms that mean contract value rises sharply by 65% among exposed buyers.

These findings suggest that buyers more vulnerable to generative AI substitution reorganize their outsourcing by consolidating into fewer, larger contracts rather than simply scaling

[18] The measure captures whether a buyer has any task vulnerable to AI substitution, which is the relevant margin if buyers reorganize production once their most automatable tasks become candidates for automation. We also considered value-weighted mean exposure and found similar results.

[19] Column (3) of Table 2 has documented a contraction at the extensive margin: the number of buyers declines significantly by about 34 percent at the mean exposure level. The estimates below establish intensive-margin adjustment among continuing buyers.

back proportionally across all margins. This consolidation pattern is consistent with buyers bundling difficult-to-automate tasks into more complex, higher-value projects discussed in Section 4 and Appendix A.[20]

The increase in total spending among continuing buyers also contrasts with the aggregate decline documented in Section 5. This reflects composition: the aggregate effect combines the intensive margin (how continuing buyers adjust) with the extensive margin (buyer exit). Column (3) of Table 2 shows that high-exposure buyers are significantly more likely to exit the platform entirely. Among those who remain, the shift toward fewer, larger contracts produces higher total spending despite reduced contract frequency.

To assess whether high- and low-exposure buyers were already on differential trajectories prior to generative AI, we estimate an event-study analog of Equation (3) in Figure 3. Event-study plots for contract counts and average contract value show flat pre-trends and sharp post-period adjustments, strengthening the interpretation that observed changes reflect responses to AI-driven capability shifts.[21]

The shift toward larger contracts is accompanied by changes in buyers' stated willingness to pay. Table 4 examines budget parameters that buyers specify when posting contracts. Column (1) shows that the maximum budget increases by 27 percent for mean-exposure buyers, while column (2) documents a 19-percent rise in minimum budgets. The increase in the ceiling exceeds that of the floor: column (3) reveals that the budget spread, that is, the difference between maximum and minimum budgets, widens significantly. The widening budget spread suggests that exposed buyers not only pay more but also exhibit greater uncertainty about appropriate compensation for the tasks they continue to outsource. This pattern is consistent with a shift toward broader, less standardized projects for which market prices are not yet well established. The increased willingness to pay a premium, combined with greater variance in stated budgets, reinforces the interpretation that buyers are reorganizing outsourcing toward tasks that are inherently more complex and difficult to price.

[20] An alternative interpretation is that firms substitute away from this platform toward other intermediaries or off-platform contracting. While we cannot directly observe off-platform activity, three observations argue against pure disintermediation. First, the contraction in outsourcing is sharply increasing in task-level automation exposure, which is difficult to reconcile with platform-specific substitution. Second, we document systematic changes in the structure of remaining projects in Section 6.2, including consolidation, higher budgets, and broader skill requirements, which align with the substitution of automatable tasks rather than simple platform exit. Third, worker-side evidence in Section 7 shows directed reskilling toward AI-complementary skills rather than diffuse reallocation, consistent with technology-driven reorganization.

[21] Appendix D.4 explores how the responses to generative AI vary across buyers and finds evidence of selective reorganization rather than uniform contraction on the demand side. Exposure to AI shifts outsourcing toward higher-value work among buyers with the technical scope to adapt, while buyers with greater scale are more likely to substitute away from outsourced production.

6.2 Task Complexity and AI-Complementary Projects

The consolidation documented above—fewer contracts at higher values—could reflect several distinct mechanisms. Buyers may simply reprice equivalent work in response to tighter labor supply, or they may fundamentally reorganize production by bundling tasks into more complex projects. To distinguish between these interpretations, we examine direct measures of project complexity observable in the transaction data.

We focus on two proxies for project scope and complexity aggregated to the buyer-week level. The first is *turnover time*, defined as the number of days between contract posting and the date a review is left, capturing the realized duration of the buyer-seller engagement. The second is *description length*, measured as the number of words in the project description. Larger and more complex projects, particularly those involving coordination across multiple skills or deliverables, typically require more detailed specifications and longer execution periods. We therefore interpret increases in these measures as evidence of greater project complexity rather than changes in pricing alone.

To isolate AI-complementary reorganization specifically, we additionally construct a measure of *AI-related work*, defined as the fraction of projects in a buyer-week that explicitly reference large language models or LLM-related tasks. This measure captures whether buyers shift toward projects involving direct engagement with generative AI tools such as deployment, customization, prompt engineering, or human oversight of automated systems. Distinguishing between general scope expansion and AI-specific integration allows us to assess whether exposed buyers are reorienting demand toward tasks that complement generative AI rather than merely scaling down pre-existing work. In parallel, we construct a measure of *interactive work*, defined as the fraction of projects in a buyer–week that require communication, client interaction, creativity, or other interpersonal skills, to assess whether reorganization also shifts outsourcing toward coordination- and relationship-intensive tasks.

Table 5 reports estimates of Equation (3) for these outcomes. The results provide consistent evidence that high-exposure buyers shift toward more complex, longer-duration projects following the introduction of ChatGPT as hypothesized in Section 4 and Appendix A. Column (1) shows that turnover time increases by 41 percent for exposed buyers. Column (2) indicates that the tasks buyers continue to outsource require more detailed specifications of requirements, deliverables, or coordination procedures. This pattern is difficult to reconcile with simple price inflation or mechanical responses to supply-side changes; longer duration and descriptions reflect increased complexity in the underlying work being contracted rather than changes in how equivalent tasks are priced.

Column (3) turns to the share of projects explicitly referencing large language models or related AI tools. The coefficient indicates that exposed buyers are significantly more likely to

post AI-related projects following ChatGPT's release. While modest in absolute terms, this effect is economically meaningful given the near-zero baseline prevalence of such references prior to November 2022. The result suggests that buyers previously reliant on automatable tasks are reorienting demand toward work that complements generative AI—projects requiring human expertise to deploy, fine-tune, or supervise AI systems rather than tasks that AI can perform autonomously.

Finally, column (4) examines the share of interactive projects, those requiring communication, client interaction, creativity, or other interpersonal skills. Although the point estimate is positive, it is not statistically significant. Together with the strong increases in project duration, description length, and AI-related content, this suggests that buyer-side reorganization operates primarily through technical complexity and AI integration rather than through increased demand for interpersonal coordination.[22]

Overall, generative AI reshapes not only how much work is outsourced, but also what kind of work remains. As routine, codifiable tasks exit the outsourced set, buyers consolidate activity into fewer, longer, and more complex projects and reorient demand toward AI-complementary tasks that require human judgment, coordination, and oversight.

6.3 AI Intensity and Outsourcing Responses

The preceding results characterize how buyers whose outsourced tasks are highly susceptible to automation consolidate remaining outsourcing as generative AI substitutes for routine production. We now examine a different margin of heterogeneity: buyers' *AI intensity*, which captures pre-existing engagement with AI-related work and reflects organizational readiness to integrate generative AI into production rather than exposure to displacement.

Table 6 reports buyer-level estimates relating AI intensity to post-ChatGPT changes in outsourcing outcomes. In contrast to the contraction and consolidation effects associated with automation exposure, AI-intensive buyers exhibit a different reallocation pattern. Column (1) shows that total outsourcing value declines for AI-intensive buyers following ChatGPT's release. At the same time, column (2) documents a significant increase in the number of contracts, while column (3) shows a sharp decline in mean contract value. These results indicate that AI-intensive buyers respond to generative AI by restructuring outsourcing into a larger number of smaller, more modular contracts.

This pattern stands in sharp contrast to the responses of buyers whose tasks are highly exposed to automation, who consolidate remaining outsourcing into fewer, higher-value projects. The opposing signs across the two buyer-level measures are economically informa-

[22]This may also reflect that interaction-intensive tasks were already concentrated among developed-country contractors or that such tasks are bundled within complex projects rather than contracted separately.

tive. Automation exposure governs whether tasks exit the outsourced set as AI substitutes for labor. In contrast, AI intensity governs how remaining outsourcing is organized once firms actively integrate AI into production. AI-intensive buyers appear to use generative AI as an internal production input and outsource complementary human inputs around automated processes, such as specialized troubleshooting, integration, evaluation, or task-specific customization, leading to finer task decomposition and smaller contract units.

The divergence in the responses suggests that generative AI reshapes outsourcing through two distinct channels. Where AI directly substitutes for human labor, outsourcing contracts and consolidates. Where firms are already positioned to use AI internally, outsourcing persists but is reorganized toward modular, complementary inputs that interface with automated systems. This distinction helps reconcile the aggregate contraction in outsourcing with the continued demand for human expertise in AI-augmented production and highlights the importance of organizational readiness in shaping firms' responses to generative AI.

7 Worker Adjustment, Reskilling, and Distribution of Opportunities

While the preceding sections document how firms reorganize outsourcing in response to advances in generative AI, an equally important margin of adjustment operates on the supply side of the market. A central advantage of our setting is that we observe realized contracts and the skills attached to those contracts over time, allowing us to study whether workers whose skills become more automatable exit the market or instead adapt by reallocating their labor toward skills that remain complementary to human expertise. Existing studies of automation typically infer worker adjustment from changes in employment across occupations or industries. In contrast, our setting allows us to observe workers' job skill portfolios at high frequency, enabling direct measurement of within-worker reskilling in response to technological change.

Guided by the task-based framework in Section 4 and Appendix A, we organize the analysis around three margins: (i) how exposed workers adjust their level of activity in the market, (ii) how the composition of their skills changes and the adoption of new skills occurs, and (iii) how earnings responds to reskilling. We then examine how these forces jointly shape the distribution of opportunities and earnings among workers.

7.1 Worker Adjustments

We examine how workers in more automatable tasks adjust their activity and earnings following the diffusion of generative AI. For worker i in week t, we estimate

$$y_{it} = \gamma_i + \lambda_{ct} + \beta \left(\text{Exposure}_i \times PostGPT_t \right) + \varepsilon_{it}, \tag{4}$$

where y_{it} denotes outcomes including the number of jobs completed, total earnings, and mean job value. Exposure_i is a time-invariant measure of worker exposure, defined as the maximum automation score across all skills utilized by worker i in tasks performed prior to November 30, 2022.[23] Worker fixed effects γ_i absorb time-invariant differences in ability, experience, and specialization, while country-time fixed effects λ_{ct} control for country-specific demand conditions and platform-wide shocks.

Table 7 examines how workers who remain active on the platform adjust along the intensive margin following the introduction of ChatGPT.[24] The results reveal a pattern of selective contraction among high-exposure workers that parallels the buyer-side reorganization documented above. Column (2) reports estimates for the log number of completed contracts per worker-week. At the mean exposure level of 0.52, this estimate translates to a 37-percent decline in the number of contracts completed, indicating a sharp contraction in demand for workers specializing in automatable tasks. Column (1) examines total earnings, measured as the log dollar value of completed contracts per worker-week. The point estimate is negative but not statistically distinguishable from zero. The imprecision of this estimate, combined with the large and precisely estimated decline in contract volume, suggests that changes in contract composition partially offset the reduction in transaction frequency. Column (3) confirms this interpretation: at the mean exposure, continuing workers earn approximately 41 percent more per contract in the post-period relative to less exposed workers.

The contrast between columns (2) and (3) is informative about the nature of worker adjustment. Rather than accepting lower compensation to maintain employment, high-exposure workers experience fewer opportunities but command higher earnings per engagement. This pattern is inconsistent with a uniform demand contraction and instead points to a shift in the composition of tasks performed by continuing workers toward higher-value, less automatable activities. The worker-side evidence thus mirrors the buyer-side consolidation

[23] As with the buyer exposure, the worker exposure measure captures whether a worker has any skill vulnerable to AI substitution, which is the relevant margin if workers decide to reorganize their skill portfolios and reallocate away from their most automatable skills. We also considered alternative aggregations, including value-weighted mean exposure and found similar results.

[24] Column (4) of Table 2 has examined the extensive margin and finds that workers with high-exposure skills are more likely to exit the platform entirely, leading to an average 35-percent reduction in the number of workers. The estimates below therefore capture intensive-margin adjustment among continuing workers.

into fewer, more complex projects documented in Section 6.1. The results from both sides of the market suggest that the aggregate contraction in platform activity reflects genuine task displacement rather than temporary frictions in matching or price adjustment. For workers, generative AI induces both displacement from routine tasks and opportunities for those able to supply skills that remain difficult to automate.

To examine whether generative AI affects the quality of work performed, we re-estimate Equation (4) using client ratings across multiple dimensions, including communication, professionalism, experience, and overall quality. These measures capture client assessments of worker performance before and after the diffusion of generative AI.

Average ratings exhibit little change following the introduction of ChatGPT. However, allowing for heterogeneity reveals a more nuanced pattern. As shown in Figure 5, quality improvements are concentrated among workers with lower pre-GPT ratings. This finding suggests that generative AI may disproportionately benefit previously underperforming workers, potentially by helping them bridge skill gaps through AI-assisted tools, improved access to information, or enhanced task execution. In contrast, workers who already performed at a high level may experience smaller marginal gains from AI assistance, consistent with diminishing returns to augmentation at the top of the performance distribution.

7.2 Skill Reallocation and Directed Reskilling

The preceding analysis shows that continuing workers perform fewer but higher-value contracts following the diffusion of generative AI. A central question is whether this adjustment reflects passive selection whereby workers simply retain their most remunerative tasks while losing lower-value work or active reallocation of the skills they supply. We address this distinction by examining how workers restructure their skill portfolios in response to exposure to generative AI.

Skill portfolio restructuring. We characterize workers' skill portfolios along four dimensions: (i) the total number of distinct skills supplied in a given week, (ii) the number of newly adopted skills not previously used by the worker, (iii) the number of pre-existing skills retained from the pre-ChatGPT period, and (iv) the average number of skills required per contract. These measures allow us to distinguish portfolio contraction from active portfolio restructuring in which displaced skills are replaced with new competencies.

Table 8 reveals pronounced restructuring among high-exposure workers. Column (1) shows that the total number of distinct skills engaged declines significantly with exposure: at the mean exposure level, workers reduce the breadth of skills applied in their work by approximately 24 percent. On its own, this pattern could reflect simple contraction.

However, columns (2) and (3) reveal substantial compositional change beneath this aggregate decline. High-exposure workers significantly increase the number of newly adopted skills while simultaneously reducing reliance on skills used prior to ChatGPT's introduction. The opposing signs indicate active replacement of skills vulnerable to automation with new competencies, rather than passive narrowing of activity.

Column (4) provides further evidence of restructuring by examining skill requirements at the contract level. High-exposure workers increasingly complete projects that demand a broader set of skills within a single engagement. This shift toward multi-skill contracts mirrors the buyer-side consolidation documented in Section 6: as buyers bundle previously separable tasks into more complex projects, workers who remain competitive must supply complementary skills jointly rather than in isolation.[25]

Columns (5) and (6) characterize the nature of the skills toward which workers reallocate. Column (5) shows that high-exposure workers are significantly more likely to engage in projects explicitly referencing large language models or AI-related tasks following ChatGPT's introduction. Column (6) shows an even larger increase in the share of projects involving interactive skills such as communication, coordination, creativity, or client-facing judgment. These results indicate that reskilling is directed toward activities that complement generative AI, either by interfacing directly with AI systems or by supplying interpersonal and contextual inputs that remain difficult to automate.

Figure 6 provides an illustrative example by visualizing transitions for workers initially specialized in Adobe Photoshop, a skill with high AI automation exposure. Flow widths are proportional to the number of workers following each transition. Adaptation in this example is highly structured. A substantial share of workers continue to supply Photoshop or closely related skills, indicating that not all activity in high-exposure domains is immediately displaced. Among those who reallocate, transitions occur primarily toward technologically proximate skills. Some workers move toward technical skills such as HTML/CSS, JavaScript, and WordPress, where generative AI assists with code generation and debugging but human input remains essential for integration and implementation. Others shift toward adjacent creative design tasks that rely on judgment and customization, or toward marketing and content-related skills that require client interaction and strategic input.

To characterize the patterns systematically, we estimate a gravity-style specification mo-

[25] A potential concern is that observed changes in skill portfolios reflect relabeling rather than genuine reskilling. This concern is mitigated in our setting because the platform maintains a stable, centrally administered skill taxonomy that workers cannot redefine. Moreover, we validate reskilling using independent evidence from project content in columns (5) and (6) of Table 8. Newly adopted skills coincide with sharp changes in the textual content of completed projects, which are difficult to reconcile with purely cosmetic tagging.

tivated by the framework in Appendix A. Appendix D.6 shows that transitions are shaped by both adoption frictions and expected payoffs. Skill-pair relatedness strongly predicts transition probability, indicating that workers move primarily along lower-friction margins. Conditional on relatedness, workers reallocate toward interactive-intensive tasks, where AI enhances productivity without requiring specialized technical retooling.

Who reskills? The preceding analysis establishes that exposed workers respond to generative AI by reallocating across skills and toward AI-complementary activities. We next examine which workers are most likely to engage in this adjustment. This distinction is important for assessing whether adaptation to generative AI is broadly shared or instead concentrated among workers with greater capacity or incentives to reskill.

We focus on two margins of skill adjustment: the acquisition of new skills and the shedding of previously supplied skills. Specifically, we define indicators for whether a worker adds at least one new skill in a given week, adds an AI-related or interaction-intensive skill, or drops at least one previously used skill. We then examine how these responses vary with worker characteristics measured prior to the diffusion of generative AI.

Table 9 reports estimates that allow reskilling responses to differ by pre-GPT experience and baseline performance. Panel A focuses on heterogeneity by experience, measured by the log number of projects completed prior to ChatGPT. The negative coefficient on *Post-GPT* $\times$ Exposure indicates that, on average, higher exposure is associated with fewer skill additions and greater skill shedding. However, the interaction with pre-GPT experience reveals a markedly different pattern for more experienced workers. Conditional on exposure, workers with more prior experience are significantly more likely to add new skills—including AI-related skills and significantly less likely to drop existing skills following the diffusion of generative AI. This pattern points to skill expansion rather than contraction among experienced workers facing automation pressure. Panel B documents analogous heterogeneity by baseline performance, measured using pre-GPT ratings. Higher-rated workers are significantly more likely to add interaction-intensive skills and substantially less likely to drop previously supplied skills after the introduction of generative AI.

These results highlight that reskilling in response to generative AI is highly uneven across workers. Among those exposed to automation, active reconfiguration of skill portfolios is concentrated among workers with stronger pre-existing market positions, i.e., those who are more experienced or higher rated. This heterogeneity provides a micro-foundation for the distributional effects examined below: generative AI reshapes not only the composition of tasks performed on the platform, but also the allocation of opportunities across workers, with adjustment and gains increasingly mediated by workers' ability to reorganize their skills in

response to automation pressure.

7.3 Earning Responses to Reskilling

Does reskilling mitigate the earnings impact of generative AI? Table 10 examines how different modes of skill adjustment mediate the relationship between exposure to generative AI and worker earnings. Across specifications, the baseline interaction between exposure and the post-GPT period is small and statistically insignificant, indicating that exposure alone does not mechanically translate into earnings losses or gains. Instead, earnings responses depend critically on how workers adjust their skill portfolios.

In contrast, active reskilling is associated with substantial earnings gains. Column (1) shows that adding new skills significantly amplifies the post-GPT exposure effect: at the mean exposure level, the addition of one new skill is associated with an approximately 21 percent increase in earnings. The gains are considerably larger when reskilling is directed toward AI-related domains. Column (2) indicates that adding AI-specific skills is associated with earnings increases of roughly 63 percent. Column (3) shows that additions of interaction-intensive skills, such as those involving communication, creativity, or client interaction, yield earnings gains of comparable magnitude.

A more informative comparison is between workers who shed skills without reorienting versus those who offset losses with new additions. Column (4) examines workers who drop previously supplied skills: earnings losses are significant when skill shedding is not accompanied by compensating additions, but are fully offset among workers who simultaneously add AI-related or interactive skills. This pattern indicates that displacement alone does not determine earnings trajectories; what matters is whether workers successfully redirect effort toward complementary domains. Column (5) confirms this interpretation: net switching intensity, which captures the balance of additions over subtractions, is positively associated with earnings.[26] In sum, generative AI does not uniformly reward or penalize exposed workers. Earnings depend on how workers respond: those who expand into AI-related and interactive domains maintain or increase earnings, while those who experience skill loss without compensating reorientation fall behind.

[26]These associations should be interpreted cautiously, as workers who successfully reskill may differ on unobservables. We view these results as suggestive evidence on the returns to reorientation rather than as causal estimates of reskilling effects.

7.4 Concentration of Market Opportunities

The preceding sections document how generative AI reshapes the volume and composition of outsourced work and how workers adjust through restructuring and reskilling. A natural question is how these responses aggregate into changes in market concentration. Generative AI can affect concentration through two distinct channels: a *within-skill* channel, where concentration rises within existing skill categories as some workers exit or lose market share to top performers, and a *cross-skill* channel, where activity reallocates toward skills that are less automatable.

The conceptual framework in Section 4 offers predictions for each margin. Within-skill concentration, by contrast, is predicted to rise unambiguously through threshold effects (consolidation raises capability requirements), selection (differential exit across the ability distribution), and differential adaptation (high-ability workers leverage AI tools and reskill more effectively). Cross-skill reallocation is ambiguous *a priori*: if high-exposure skills were more concentrated than average before AI, their contraction would reduce aggregate concentration; the reverse would increase it. We now examine whether concentration rises in practice and, critically, through which channels.

To understand the mechanisms driving concentration changes, we decompose aggregate concentration following the framework developed in Appendix A. Formally, let s_{it} denote skill i's share of platform earnings or contracts at time t, and C_{it} denote concentration (the fraction of earnings or contracts by the top five workers) in skill i at time t. Aggregate concentration is $\bar{C}_t = \sum_i s_{it} C_{it}$. Changes in aggregate concentration decompose as:

$$\Delta \bar{C} = \underbrace{\sum_i \bar{s}_i \Delta C_i}_{\text{within-skill}} + \underbrace{\sum_i \bar{C}_i \Delta s_i}_{\text{cross-skill}} + \text{interaction}, \tag{5}$$

where $\bar{s}_i$ and $\bar{C}_i$ denote baseline (pre-GPT) values. The within-skill component captures concentration changes within skills weighted by their baseline importance; the cross-field component captures reallocation of activity across skills weighted by their baseline concentration levels.

We further decompose the within-skill component to distinguish the role of worker turnover from rising inequality among continuing workers. The *participation* margin captures how concentration changes due to the entry and exit of workers. The *within-active* margin captures how concentration changes among workers who were active before GPT and remain active afterward. Let C_{it}^{pre} denote the top-5 share in skill i at time t computed only among

workers who were active in that skill prior to GPT. Then:

$$\Delta C_i = \underbrace{(C_{it} - C_{it}^{\text{pregpt}})}_{\text{participation}} + \underbrace{(C_{it}^{\text{pregpt}} - C_{i,\text{baseline}})}_{\text{within-active}}. \tag{6}$$

The within-active component is positive if concentration among continuing pre-GPT workers exceeds the baseline, indicating that top performers among stayers capture a larger share over time. As discussed in Appendix A, the sign of the participation margin provides a diagnostic for the nature of selection: a positive coefficient indicates exit concentrated among low performers, while a negative coefficient suggests that some top performers also exit due to depreciated task-specific human capital or superior outside options.

We estimate the effect of AI exposure on each concentration component using Equation (1) from Section 5, including the top-5 share (within-skill concentration), the skill's contribution to cross-skill concentration, the participation component, or the within-active component. Table 11 reports the decomposition results. Panel A examines concentration within each skill category. The positive and significant coefficients imply that at the mean AI automation exposure level, within-skill concentration increases by approximately 27 percentage points for earnings and 26 percentage points for projects following the diffusion of generative AI, a substantial increase in the concentration of market opportunities.

Panel B examines whether platform activity shifts toward skills that were already concentrated. The coefficients are negative and economically small, indicating that high-exposure skills, which contract following AI introduction as documented in Section 5, were slightly more concentrated than average before GPT, so their contraction marginally reduces aggregate concentration through the cross-skill channel. However, the magnitude is negligible: cross-skill reallocation contributes approximately one percentage point to the change in aggregate concentration. Rising platform-wide concentration is driven primarily by what happens *within* skill categories.

Panel C decomposes the within-skill effect into participation and within-active margins. This finding is initially surprising. Under the simplest version of the theoretical framework where low-ability workers exit while high-ability workers remain, we would expect a positive participation effect, as the departure of low performers mechanically raises the share captured by top performers. The negative coefficient implies the opposite: turnover patterns actually reduce measured concentration in high-exposure skills, suggesting that displacement is not confined to low-performing contractors.

However, this result is consistent with the reskilling patterns documented in Section 7.2. Top performers in high-exposure skills are more likely to successfully reskill toward AI-complementary tasks, effectively exiting their original skill category while entering new ones.

The negative participation effect thus reflects two forces working in the same direction: (i) some high performers exit high-exposure skills because they successfully transition to AI-complementary work; and (ii) new entrants or workers expanding from other skills dilute concentration in high-exposure categories. Both mechanisms reduce measured concentration relative to what would prevail if only pre-GPT workers in that skill were counted.

The within-active coefficient is positive and significant, suggesting that among workers who were active in a skill before GPT and continue participating in that same skill, top performers capture an increasing share of opportunities. This finding supports the framework's prediction that AI-driven reorganization favors high-ability workers through selection effects and rising capability thresholds.

Panel D translates the skill-level coefficients into implied effects on platform-wide concentration. Within-skill concentration contributes approximately 27 percentage points to the aggregate increase in earnings concentration. The cross-skill component reduces concentration by approximately 1 percentage point, indicating a positive correlation between the extent of automation exposure across skills and pre-GPT concentration levels. The total implied effect on market concentration is approximately 26 percentage points for earnings and 25 percentage points for projects.

These findings demonstrate that generative AI reshapes not only the volume but also the distribution of opportunities in developing-country digital labor markets. Demand for highly exposed skills contracts sharply, and within the reduced set of remaining tasks, earnings and projects become increasingly concentrated among top performers. The workers best positioned to adapt are also the most likely to reskill into AI-complementary activities, accelerating exit from exposed skill categories. Because the largest contractions occur in skills that were already relatively concentrated, cross-skill activity shares change little; instead, rising concentration emerges primarily within skills. Skill categories thus remain stable organizational units, even as inequality intensifies among workers competing within the same tasks. This pattern highlights a central mechanism of AI-driven displacement in outsourcing markets: task-specific human capital may inhibit worker transition, while broader, more general skills facilitate adaptation and insulate top performers from displacement.

8 Conclusion

Generative artificial intelligence is reshaping not only how work is performed, but where it is produced. This paper studies the early reorganization of international trade in digital services following the introduction of generative AI, using transaction-level data from a global online labor marketplace. By observing realized contracts rather than intended demand and jointly

analyzing buyer reorganization and worker adjustment within a single market, we provide a unified account of how a general-purpose technology alters the organization and geography of production. We document large effects operating through both sides of the market: a 35 percent decline in developing-country outsourcing, consolidation of buyer demand into complex engagements, structured worker reskilling with stark heterogeneity by ability, and rising concentration of opportunities among top performers. These patterns are consistent with a framework in which AI compresses the cross-border wage arbitrage that previously sustained developing-country comparative advantage in routine cognitive tasks.

Three features of our findings warrant emphasis. First, effects emerged quickly within two years of ChatGPT's release, suggesting adjustment to generative AI may be faster than historical experience with trade or automation shocks would predict. Second, distributional consequences are severe: workers in developing countries face both aggregate displacement and rising inequality, with the capacity to adapt concentrated among those already best positioned. Third, the international dimension is central. What appears as productivity gains from the perspective of developed-country firms manifests as displacement in the developing-country labor markets that supply these services.

Our findings carry policy implications at multiple levels. For developing countries, digital services exports, a sector that offered a pathway to growth without the capital requirements of manufacturing, may be more vulnerable to disruption than understood. The affected workers are educated professionals who invested in precisely the cognitive skills that development strategies encouraged. For developed countries, the impact is more ambiguous: buyers benefit from lower costs, but the labor-market effects of automation and reduced offshoring remain uncertain. At the international level, the countries driving AI innovation are not those bearing the largest adjustment costs, an asymmetry that raises questions about whether existing frameworks for managing technological change are adequate.

Several directions for future research emerge from our analysis. We study a single platform during a period of rapid capability growth; whether effects generalize to traditional offshoring arrangements or domestic labor markets, and whether short-run displacement gives way to new forms of human-AI complementarity, remain open questions. We observe market-level adjustment but not the micro-foundations of individual decisions: why some workers successfully reskill while others do not.

The rise of generative AI may represent an inflection point in global production. For two decades, declining communication costs enabled unprecedented unbundling of tasks across borders. Our evidence suggests AI may partially reverse this unbundling, not by reshoring tasks to developed countries, but by eliminating the need to assign them to human workers at all. Whether this represents temporary disruption or fundamental transformation remains

uncertain. What is clear is that the effects are already large, already unequal, and already reshaping the markets where digital work is bought and sold.

References

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, 1998, *113* (4), 1055–1089.

_ **and David Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171.

_ **and Pascual Restrepo**, "Artificial intelligence, automation, and work," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 197–236.

_ **and** _ , "Artificial Intelligence and Jobs," *Journal of Economic Perspectives*, 2020, *34* (3), 30–50.

_ **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Agrawal, Ajay, John J. Horton, Nicola Lacetera, and Elizabeth Lyons, "Digitization and the Contract Labor Market: A Research Agenda," *Management Science*, 2016, *62* (7), 1953–1972.

Alekseeva, Liudmila, José Azar, Mireia Gine, Sampsa Samila, and Bledi Taska, "The demand for AI skills in the labor market," *Labour Economics*, 2021, *71*, 102002.

Antràs, Pol, *Global Production: Firms, Contracts, and Trade Structure*, Princeton University Press, 2015.

_ **and Davin Chor**, "Organizing the Global Value Chain," *Econometrica*, 2013, *81* (6), 2127–2204.

Artuc, Erhan, Paulo Bastos, and Bob Rijkers, "Robots, tasks, and trade," *Journal of International Economics*, 2023, *145*, 103828.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333.

Baldwin, Richard, *The Globotics Upheaval: Globalization, Robotics, and the Future of Work*, Oxford University Press, 2019.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

_ , Tom Mitchell, and Daniel Rock, "What can machines learn, and what does it mean for occupations and the economy?," *AEA papers and proceedings*, 2018, *108*, 43–47.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani, "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality," *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, 2023, (24-013).

DeStefano, Timothy and Jonathan Timmis, "Robots and export quality," *Journal of Development Economics*, 2024, *168*, 103248.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor Market Impact Potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308.

Faber, Marius, "Robots and reshoring: Evidence from Mexican labor markets," *Journal of International Economics*, 2020, *127*, 103384.

Felten, Edward W, Manav Raj, and Robert Seamans, "The occupational impact of artificial intelligence: Labor, skills, and polarization," *NYU Stern School of Business*, 2019.

Grossman, Gene M and Esteban Rossi-Hansberg, "Trading tasks: A simple theory of offshoring," *American Economic Review*, 2008, *98* (5), 1978–1997.

Grossman, Gene M. and Esteban Rossi-Hansberg, "Task Trade between Similar Countries," *Econometrica*, 2012, *80* (2), 593–629.

Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K. Troy Dario Amodei, Jared Kaplan, Jack Clark, and Deep Ganguli, "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," 2025.

Horton, John J., "Online Labor Markets," in Amin Saberi, ed., *Internet and Network Economics*, Vol. 6484 Springer 2010, pp. 515–522.

Hui, Xuan, Oren Reshef, and Luofeng Zhou, "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," *Organization Science*, 2024.

Lou, Bowen, Hongshen Sun, and Tianshu Sun, "Gpts and labor markets in the developing economy: Evidence from china," *Available at SSRN 4426461*, 2023.

Noy, Shakked and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, 2023, *381* (6654), 187–192.

Teutloff, Ole, Johanna Einsiedler, Otto Kässi, Fabian Braesemann, Pamela Mishkin, and R. Maria del Rio-Chanona, "Winners and Losers of Generative AI: Early Evidence of Shifts in Freelancer Demand," *Journal of Economic Behavior & Organization*, 2025, *235*, 106845.

Figures and Tables

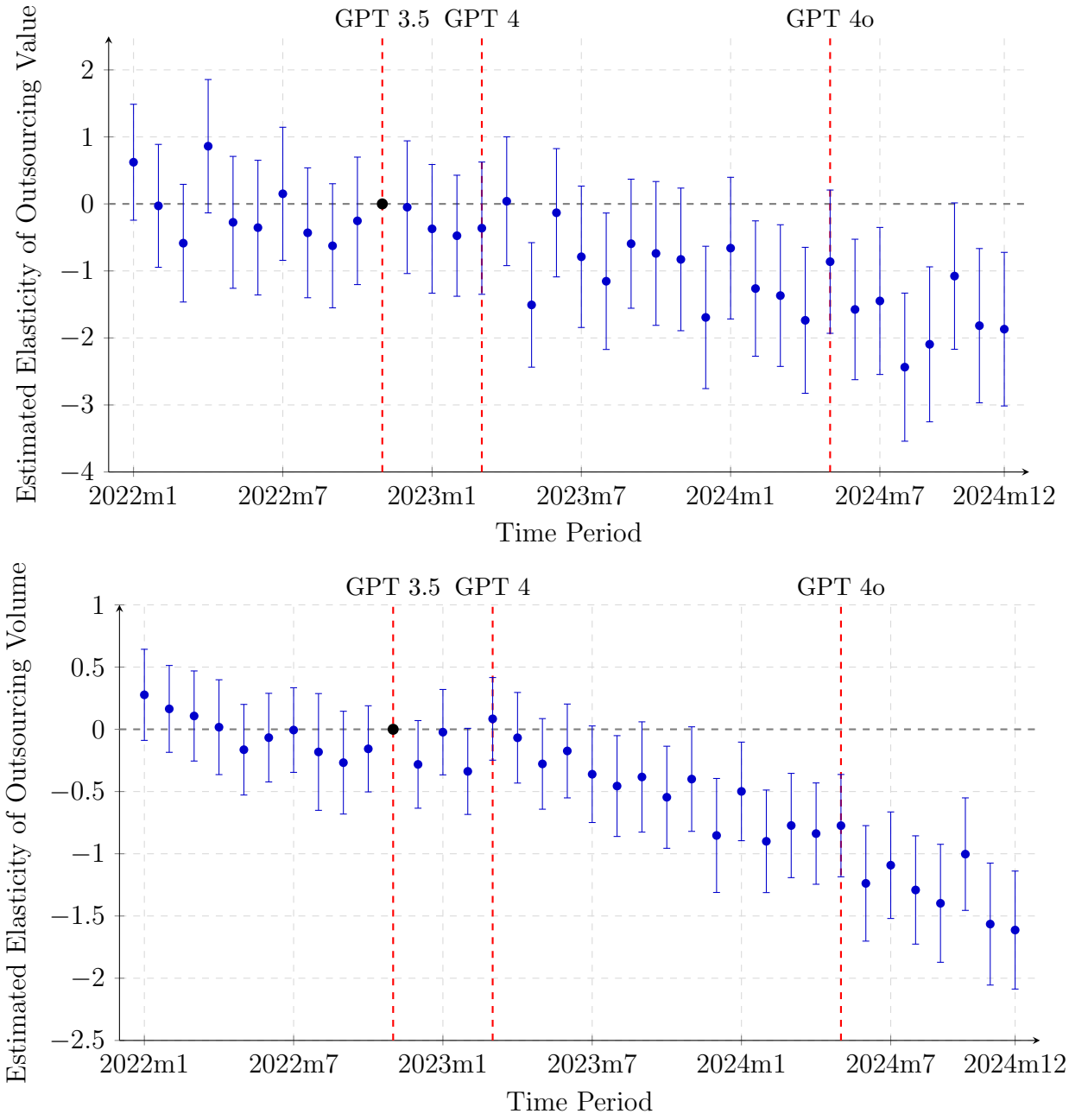


Figure 1: Event Study: Dynamic Effects of Generative AI on Outsourcing

Notes. These figures present event study estimates of the dynamic effects of generative AI model releases on international outsourcing. Each point represents the coefficient estimate for a given month relative to the baseline period, with 95% confidence intervals shown. Vertical dashed red lines indicate the release dates of GPT-3.5 (November 30, 2022), GPT-4 (March 14, 2023), and GPT-4o (May 13, 2024). The horizontal dashed gray line indicates zero effect. Standard errors are clustered at the skill level. The top panel shows the elasticity of total outsourcing value, while the bottom panel shows the elasticity of the number of contracts.

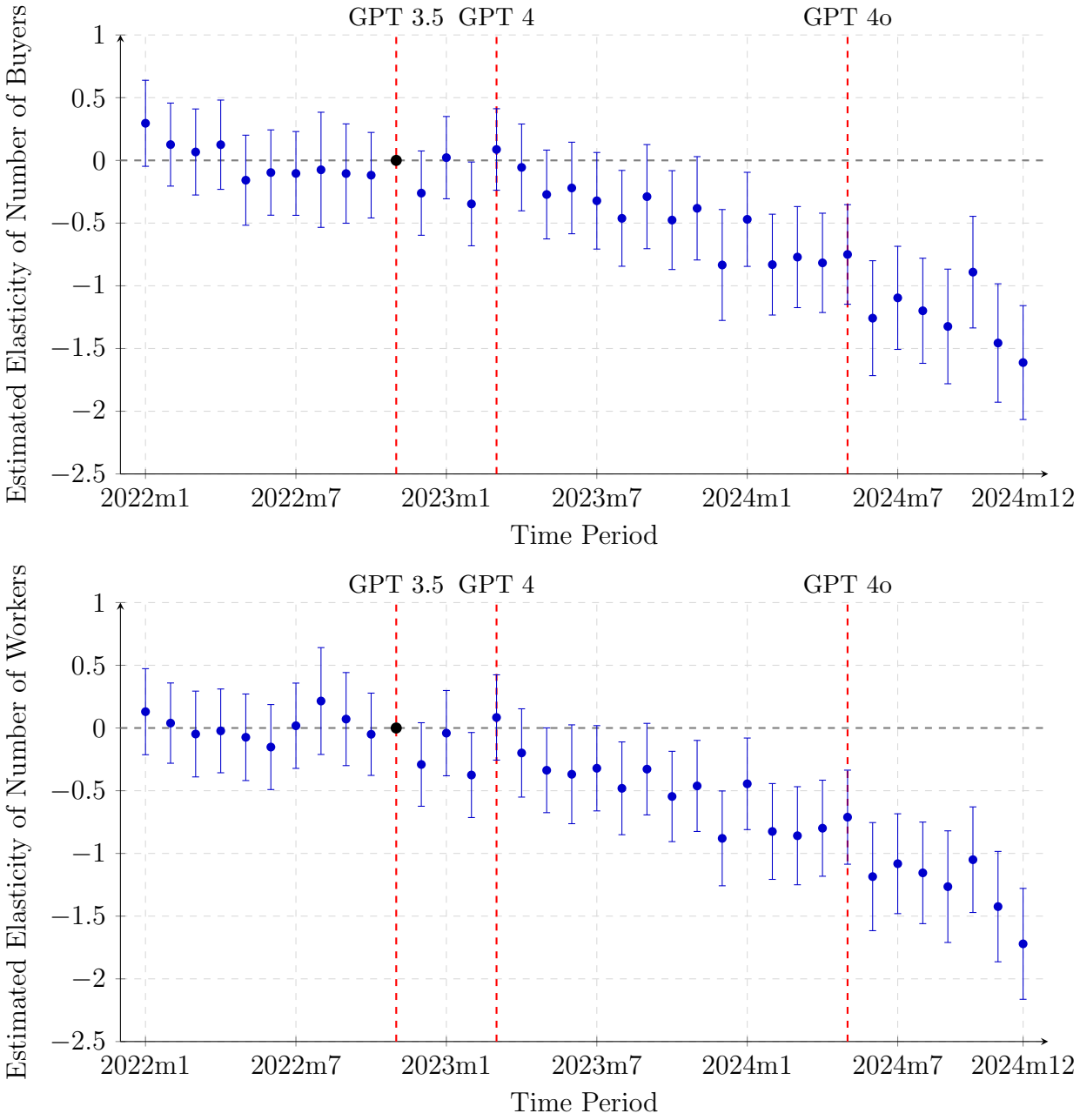


Figure 2: Event Study: Dynamic Effects of Generative AI on Buyers and Workers

Notes. These figures present event study estimates of the dynamic effects of generative AI model releases on international outsourcing. Each point represents the coefficient estimate for a given month relative to the baseline period, with 95% confidence intervals shown. Vertical dashed red lines indicate the release dates of GPT-3.5 (November 30, 2022), GPT-4 (March 14, 2023), and GPT-4o (May 13, 2024). The horizontal dashed gray line indicates zero effect. Standard errors are clustered at the skill level. The top panel shows the elasticity of the number of buyers, while the bottom panel shows the elasticity of the number of workers.

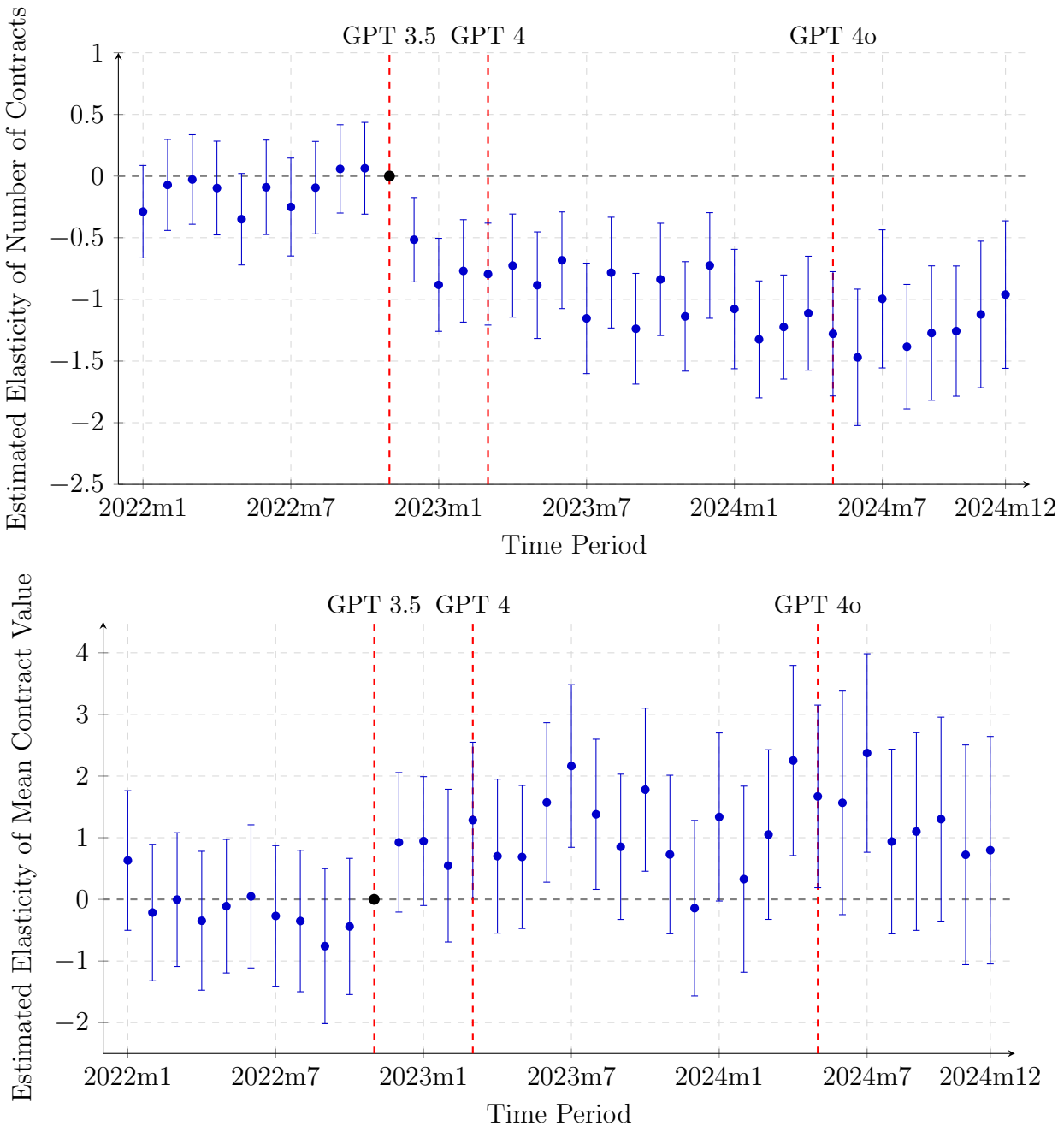


Figure 3: Event Study: Dynamic Effects of Generative AI on Buyer-Level Contract Volume and Average Contract Value

Notes. These figures present event study estimates of the dynamic effects of generative AI model releases on international outsourcing. Each point represents the coefficient estimate for a given month relative to the baseline period, with 95% confidence intervals shown. Vertical dashed red lines indicate the release dates of GPT-3.5 (November 30, 2022), GPT-4 (March 14, 2023), and GPT-4o (May 13, 2024). The horizontal dashed gray line indicates zero effect. Standard errors are clustered at the skill level. The top panel shows the elasticity of the buyers' number of contracts, while the bottom panel shows the elasticity of the buyers' average contract value.

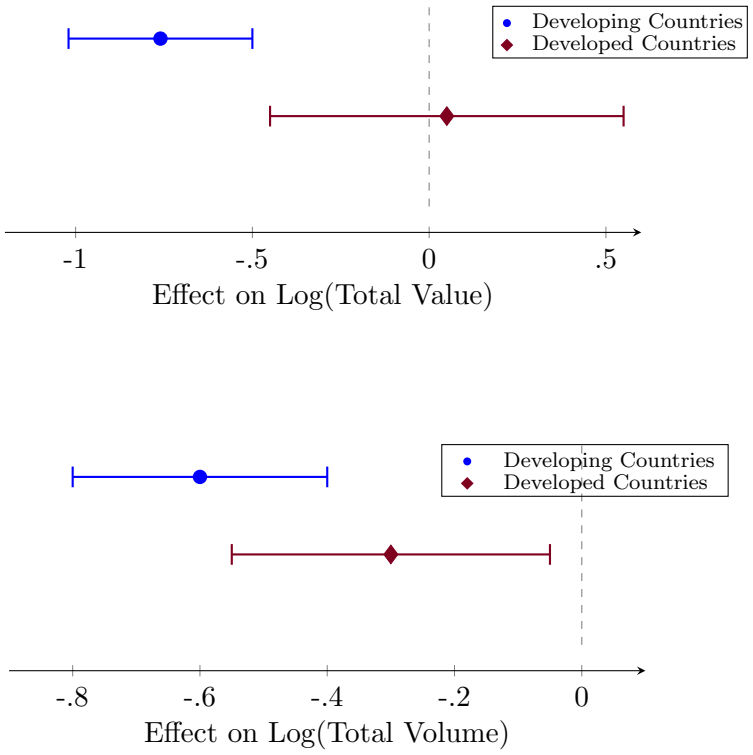


Figure 4: Aggregate Effects: Developing vs Developed Countries

Notes: The figure compares the estimated coefficients from Equation (1) for developed and developing economies, along with their 95% confidence intervals. The top panel shows the impact of generative AI on the total value of contracts outsourced to each group of countries, while the bottom panel illustrates the effect on the volume of contracts outsourced.

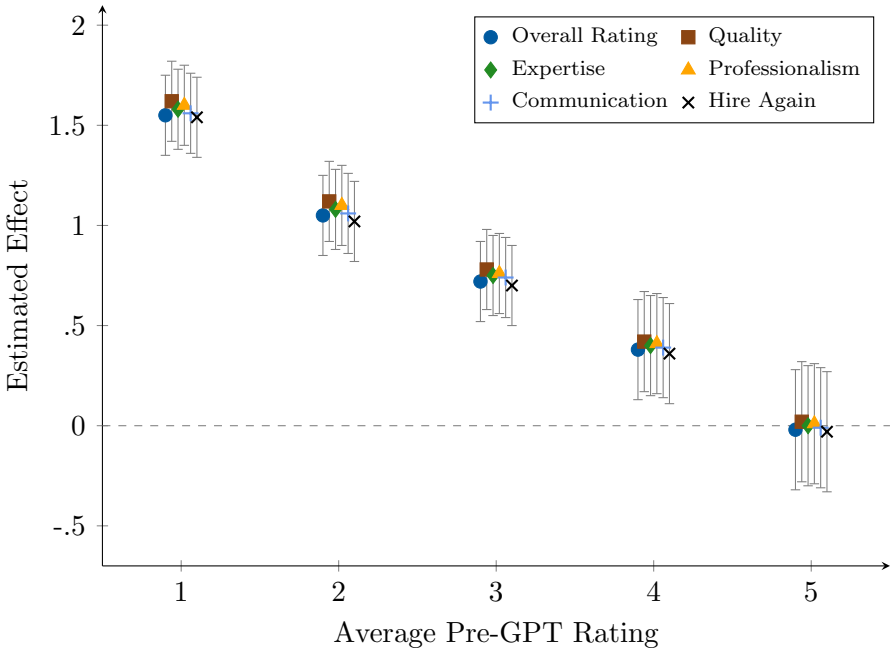


Figure 5: The Heterogeneous Effects of Generative AI across Worker Quality

Notes: The figure plots the estimated effects of generative AI on contract volume across workers based on the levels of their average ratings before the introduction of ChatGPT, along with their 95% confidence intervals. Ratings include overall rating and ratings on quality, expertise, professionalism, communication, and the likelihood to hire again.

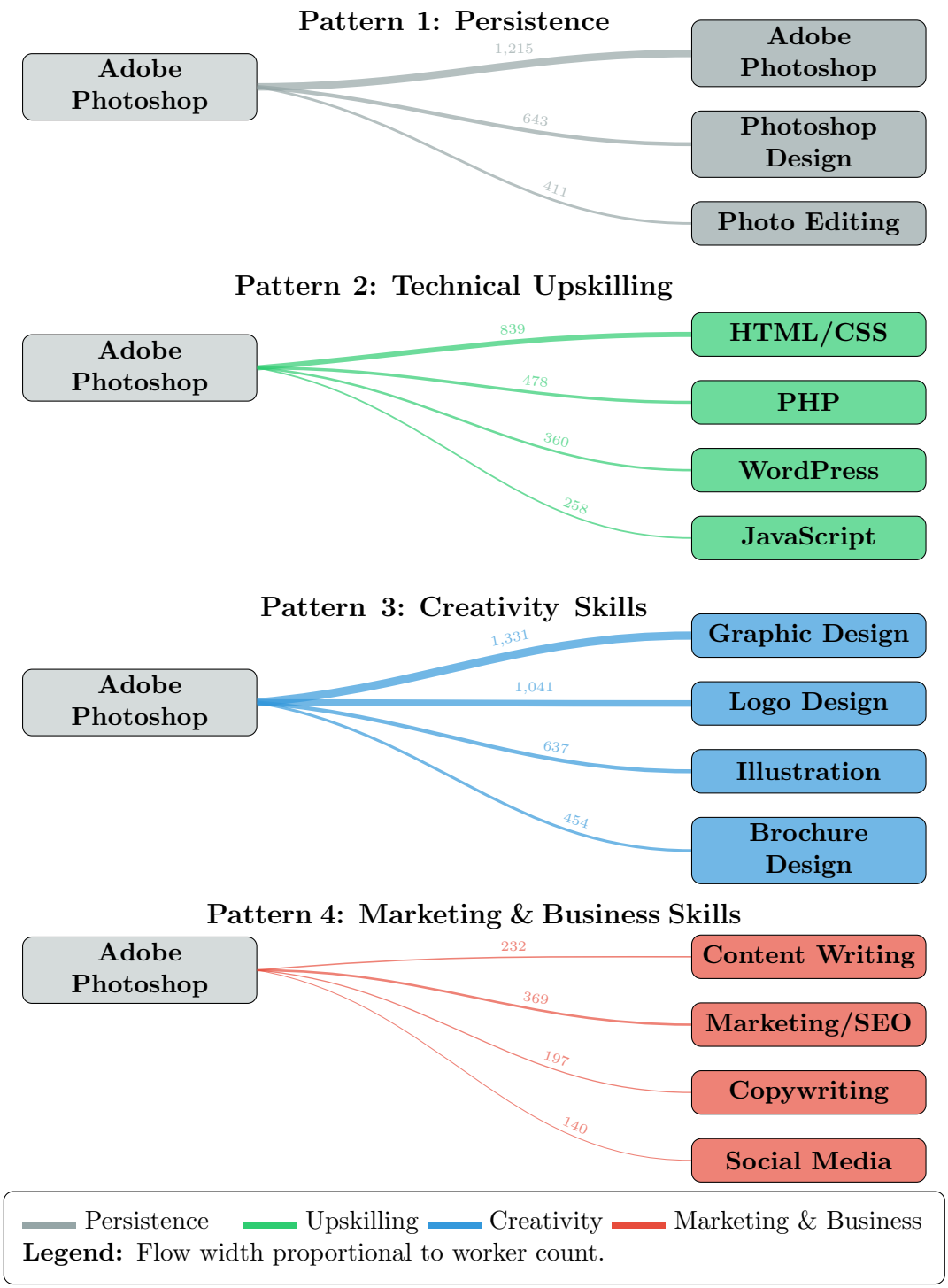


Figure 6: Worker Adaptation Strategies from High-AI-Exposure Skills

Notes. This figure maps skill transitions from Adobe Photoshop (a skill with high AI automation exposure) following ChatGPT's November 30, 2022 release, organized by adaptation strategy. *Pattern 1 (Persistence):* 2,269 workers remain in Photoshop-related skills despite high AI substitution risk. *Pattern 2 (Technical Upskilling):* 1,935 workers transition to complementary technical skills (HTML, PHP, JavaScript) that involve system architecture and logic—areas where AI serves as a tool rather than a substitute. *Pattern 3 (Creativity Skills):* 3,463 workers shift to related design skills (Graphic Design, Logo Design). *Pattern 4 (Marketing & Business Skills):* 938 workers move to marketing and content strategy tasks requiring client interaction, business judgment, and creative strategy. These patterns reveal that while some workers persist within high-exposure domains, a substantial portion (Patterns 2-4) reposition toward technical, creative, or client-facing tasks.

Table 1: Most and Least Generative AI Automatable Skills

Most Automatable Skills			Least Automatable Skills		
Rank	**Skill**	**Score**	**Rank**	**Skill**	**Score**
1	Google Search	0.948	2718	Anthropology	0.314
2	Software Engineering	0.943	2719	Attic Access Ladders Making	0.314
3	English Spelling	0.930	2720	Christmas Planning	0.312
4	Programming	0.929	2721	Genealogy	0.308
5	English Grammar	0.928	2722	Abnormal Psychology	0.306
6	Coding	0.925	2723	Forth (Programming Language)	0.306
7	Word Processing	0.924	2724	Feng Shui	0.302
8	English (UK) Translation	0.921	2725	Alto Flute	0.295
9	HTML	0.916	2726	Palm	0.295
10	Adobe Photoshop	0.914	2727	Soldering	0.287
11	Android Development	0.913	2728	COBOL	0.282
12	Proofreading	0.905	2729	ASHSP	0.279
13	CMS	0.905	2730	MMORPG	0.274
14	Mathematics	0.905	2731	Building Certification	0.258
15	Writing	0.905	2732	Symbian	0.254
16	English (US) Translation	0.904	2733	Asbestos Removal	0.250
17	Typing	0.904	2734	Papiamento	0.244
18	Google Chrome	0.903	2735	Sculpturing	0.233
19	ChatGPT Usage	0.902	2736	Google Buzz	0.225
20	Web Development	0.902	2737	Fire Safety Planning	0.224
21	Copy Typing	0.902	2738	ECPay	0.200
22	Firefox	0.901	2739	Fashion Modeling	0.192
23	Photoshop	0.899	2740	Google Wave	0.191
24	Web Applications	0.896	2741	MySpace	0.144
25	Drawing / Artist	0.895	2742	Flash Mob	0.132

Notes. This table reports the 25 skills with the highest and lowest exposure to generative AI automation among the 2,700 platform-defined skill categories. Exposure scores are constructed by eliciting assessments from GPT-3.5 of the extent to which generative AI can perform the tasks associated with each skill with minimal human input. Scores are averaged over 100 independent, stateless prompt iterations using fixed decoding parameters and normalized to lie between 0 and 1. Higher values indicate greater automation exposure. Scores should be interpreted in relative rather than absolute terms. Skill labels correspond to the platform's fixed taxonomy and do not necessarily map one-to-one to standard occupational classifications.

Table 2: Aggregate Effects on International Outsourcing

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Buyers)	(4) Log(Workers)
Post-GPT × Exposure	-0.820***	-0.608***	-0.601***	-0.612***
	(0.147)	(0.089)	(0.085)	(0.079)
Constant	6.652***	1.491***	1.457***	1.409***
	(0.071)	(0.043)	(0.041)	(0.038)
Skill FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
R^2	0.637	0.887	0.892	0.884
Observations	88,800	90,045	90,045	90,045

Notes. The dependent variables are measured at the skill-week level. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a skill-level exposure measure capturing GPT's assessed ability to substitute for workers in that skill, normalized to the unit interval. All specifications include skill fixed effects and week fixed effects. Standard errors, clustered at the skill level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Continuing Buyers' Outsourcing Margins

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Mean Value)
Post-GPT × Exposure	1.095***	-0.360***	1.454***
	(0.217)	(0.044)	(0.212)
Constant	4.232***	0.216***	4.010***
	(0.083)	(0.017)	(0.081)
Buyer FE	Yes	Yes	Yes
Source Country × Week FE	Yes	Yes	Yes
R^2	0.581	0.354	0.593
Observations	92,820	94,910	92,820

Notes. The dependent variables are measured at the buyer-week level. *Value* is the sum of the dollar value of all completed contracts posted by a buyer in a given week, converted from local currency using contemporaneous exchange rates. *Contracts* is the number of completed contracts posted by a buyer in a given week. *Mean Value* is the average value of completed contracts. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a buyer-level exposure measure defined as the maximum automation exposure to generative AI across all contracts posted by the buyer prior to the introduction of ChatGPT and is normalized to the unit interval. All specifications include buyer fixed effects and source country-by-week fixed effects. Standard errors, clustered at the buyer level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Changes in Buyers' Willingness to Pay

	(1) Log(Max Budget)	(2) Log(Min Budget)	(3) Log(Budget Spread)
Post-GPT × Exposure	0.789***	0.564***	0.980***
	(0.203)	(0.208)	(0.303)
Constant	4.034***	3.315***	3.570***
	(0.078)	(0.079)	(0.120)
Buyer FE	Yes	Yes	Yes
Source Country × Time FE	Yes	Yes	Yes
R^2	0.608	0.644	0.632
Observations	94,657	94,779	48,164

Notes. The dependent variables are measured at the buyer-week level. *Max Budget* and *Min Budget* are the average maximum and minimum budgets of contracts posted by a buyer in a given week, converted from local currency using contemporaneous exchange rates. *Budget Spread* is the average difference between maximum and minimum budgets in contracts posted by a buyer in a given week. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a buyer-level exposure measure defined as the maximum automation exposure to generative AI across all contracts posted by the buyer prior to the introduction of ChatGPT and is normalized to the unit interval. All specifications include buyer fixed effects and source country-by-week fixed effects. Standard errors, clustered at the buyer level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Task Reorganization: Project Complexity and Interactivity

	(1) Log(Turnover Time)	(2) Log(Description)	(3) AI Share	(4) Interactive Share
Post-GPT × Exposure	1.196***	0.393**	0.023**	0.090
	(0.301)	(0.190)	(0.009)	(0.080)
Constant	1.597***	3.246***	−0.006	0.241***
	(0.115)	(0.073)	(0.004)	(0.030)
Buyer FE	Yes	Yes	Yes	Yes
Source Country × Week FE	Yes	Yes	Yes	Yes
R^2	0.521	0.519	0.331	0.425
Observations	70,804	94,910	94,910	94,910

Notes: The dependent variables are measured at the buyer-week. *Turnover Time* is the number of days between project posting and the date a review is left. *Description Length* is measured as the number of words in the project description. *AI Share* is the fraction of projects in a buyer-week that reference large language models or LLM-related tasks. *Interactive Share* is the fraction of projects in a buyer-week that require communication, interactions, and other interpersonal skills. *Post-GPT* is an indicator equal to one for weeks after November 30, 2022. *Exposure* is a buyer-level exposure measure defined as the maximum automation exposure to generative AI across all contracts posted by the buyer prior to the introduction of ChatGPT and is normalized to the unit interval. All specifications include buyer fixed effects and source country-by-week fixed effects. Standard errors are clustered at the buyer level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: AI Intensity and Outsourcing Margins

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Mean Value)
Post-GPT × AI Intensity	−0.481***	0.060***	−0.527***
	(0.140)	(0.019)	(0.138)
Constant	4.654***	0.079***	4.569***
	(0.001)	(0.000)	(0.001)
Buyer FE	Yes	Yes	Yes
Source Country × Week FE	Yes	Yes	Yes
R^2	0.581	0.354	0.593
Observations	92,820	94,910	92,820

Notes. The dependent variables are measured at the buyer-week level. *Value* is the sum of the dollar value of all completed contracts posted by a buyer in a given week, converted from local currency using contemporaneous exchange rates. *Contracts* is the number of completed contracts posted by a buyer in a given week. *Mean Value* is the average value of completed contracts. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *AI Intensity* is the share of projects involving AI-related tasks sourced by the buyer prior to ChatGPT's release and is normalized to the unit interval. All specifications include buyer fixed effects and source country-by-week fixed effects. Standard errors, clustered at the buyer level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Worker-Level Contract Adjustments

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Mean Value)
Post-GPT × Exposure	-0.190	-0.905***	0.663***
	(0.277)	(0.130)	(0.243)
Constant	5.185***	0.803***	4.405***
	(0.141)	(0.066)	(0.124)
Individual FE	Yes	Yes	Yes
Destination Country × Time FE	Yes	Yes	Yes
R^2	0.432	0.445	0.448
Observations	128,102	131,061	128,102

Notes. The dependent variables are measured at the worker-week level for continuing workers. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the number of completed contracts in a given week. *Mean Value* is the average value of completed contracts. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a worker-level exposure measure defined as the maximum exposure to GPT across all contracts performed by the worker prior to the introduction of ChatGPT and is normalized to the unit interval. All specifications include individual FE and destination country-by-week FE. Standard errors, clustered at the individual level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Worker-Level Skill Reallocations

	Skill Portfolio			Task Composition		
	Log(Total)	Log(New)	Log(Existing)	Log(Skills/Contract)	AI Share	Interact. Share
	(1)	(2)	(3)	(4)	(5)	(6)
Post-GPT × Exposure	−0.527***	0.513***	−0.397***	0.115*	0.122**	0.354***
	(0.096)	(0.160)	(0.136)	(0.061)	(0.051)	(0.090)
Constant	1.925***	0.522***	1.694***	1.439***	0.196***	0.160***
	(0.049)	(0.063)	(0.075)	(0.031)	(0.026)	(0.046)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest. Country × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.298	0.365	0.371	0.220	0.599	0.207
Observations	131,061	58,133	118,724	131,061	131,061	131,061

Notes. The unit of observation is the worker–week for continuing workers. *Skill Count* is the number of distinct skills in which a worker completes at least one contract in a given week. *New Skills* counts skills used for the first time in a given week. *Existing Skills* counts skills previously used by the worker. *Skills/Contract* is the average number of skills associated with completed contracts. *AI Share* is the fraction of projects explicitly referencing large language models or AI-related tasks. *Interactive Share* is the fraction of projects requiring communication, coordination, creativity, or other interpersonal skills. *Post-GPT* equals one for weeks after November 30, 2022. *Exposure* is a worker-level pre-GPT exposure to generative AI, normalized to $[0, 1]$. All specifications include individual fixed effects, time fixed effects, and country-by-week fixed effects. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Reskilling: Heterogeneity by Experience and Rating

	(1) Add Skill	(2) Add AI Skill	(3) Add Int. Skill	(4) Drop Skill
Panel A: Heterogeneity by Experience				
Post-GPT × Exposure	−0.407***	−0.364***	−0.330***	0.867***
	(0.102)	(0.055)	(0.110)	(0.096)
× ln(Pre-GPT Projects)	0.022***	0.031***	−0.009	−0.098***
	(0.004)	(0.002)	(0.005)	(0.004)
Panel B: Heterogeneity by Rating				
Post-GPT × Exposure	−0.070	0.174**	−0.584***	−0.139
	(0.105)	(0.070)	(0.147)	(0.117)
× Pre-GPT Rating	0.001	−0.009	0.023*	−0.107***
	(0.011)	(0.010)	(0.013)	(0.011)
Individual FE	Yes	Yes	Yes	Yes
Destination Country × Time FE	Yes	Yes	Yes	Yes
R^2	0.273	0.177	0.236	0.301
Observations	130,595	130,595	130,595	130,595

Notes. The unit of observation is the worker–week. *Add Skill* equals one if a worker performs at least one task in a new skill not previously supplied. *Add AI Skill* and *Add Int. Skill* indicate additions of AI-related and interaction-intensive skills, respectively. *Drop Skill* equals one if a worker ceases to supply at least one previously used skill. *Post-GPT* equals one for weeks after November 30, 2022. *Exposure* is a worker-level pre-GPT exposure to generative AI, normalized to $[0, 1]$. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Earnings Impact of Reskilling by Skill Type

	Dependent Variable: Log(Value)				
	(1)	(2)	(3)	(4)	(5)
Post-GPT × Exposure	−0.261	−0.106	−0.642	0.136	0.005
	(0.273)	(0.426)	(1.410)	(0.280)	(0.274)
× Skill Additions	0.374***				
	(0.044)				
× AI Skill Additions		0.934***			
		(0.302)			
× Interactive Skill Additions			0.498**		
			(0.217)		
× Skill Reductions				−0.302***	
				(0.045)	
× Switching Intensity					0.215***
					(0.026)
Individual FE	Yes	Yes	Yes	Yes	Yes
Destination Country × Time FE	Yes	Yes	Yes	Yes	Yes
R^2	0.428	0.422	0.428	0.427	0.428
Observations	124,253	92,193	71,643	124,253	124,253

Notes: The dependent variable *Log(Value)* is the log total dollar value of completed contracts per worker-week, converted from local currency using contemporaneous exchange rates. *Post-GPT* is an indicator equal to one for weeks after November 30, 2022. *Exposure* is a worker-level measure of pre-GPT exposure to generative AI, normalized to $[0, 1]$. *Skill Additions* counts newly added skills. *AI Skill Additions* counts newly added AI-related skills (e.g., machine learning, AI integration, chatbot development). *Interactive Skill Additions* counts newly added interaction-related skills, such as those involving communication or client interaction. *Skill Reductions* measures reductions in pre-GPT skills. *Switching Intensity* measures the net change in the number of active skills relative to the pre-GPT period. All skill measures are normalized by the average number of skills in the pre- and post-GPT periods. All specifications include individual fixed effects and destination country-by-week fixed effects. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Concentration Decomposition: Summary of Mechanisms

	Earnings	Projects
A. Does concentration rise in high-exposure skills?		
Within-skill concentration	0.574***	0.545***
	(0.061)	(0.061)
B. Does activity shift across skills?		
Cross-skill contribution	−0.062	−0.085**
	(0.042)	(0.041)
C. What drives within-skill concentration?		
Participation (exit/entry)	−0.236***	−0.476***
	(0.056)	(0.071)
Within-active (stayers)	0.143***	0.348***
	(0.018)	(0.042)
D. Implied aggregate effect		
Within-field ($\beta \times \bar{E}$)	0.275	0.261
Cross-field ($\beta \times \bar{E}$)	−0.01	−0.01
Total concentration	**0.265**	**0.251**

Notes: Each entry reports the coefficient on skill-level generative AI exposure interacted with a post-GPT indicator from a skill–week regression. Panel A reports estimates from regressions in which the dependent variable is within-skill concentration, measured as the top-five share of earnings or projects within a skill–week. Panel B reports estimates from regressions where the dependent variable is the cross-skill component of aggregate concentration, constructed by weighting skill-level concentration measures by pre-GPT earnings shares. Panel C reports estimates from regressions that further decompose within-skill concentration into a participation component, capturing changes driven by entry and exit of workers within skills, and a within-active component, capturing changes in the distribution of outcomes among continuing workers. Panel D reports the implied contribution of within- and cross-skill channels to aggregate concentration, computed as the product of the estimated coefficients and mean skill exposure in the sample. All specifications include skill and week fixed effects. Standard errors, reported in parentheses, are clustered at the skill level. The participation and within-active estimates in Panel C are based on different subsamples and therefore do not mechanically sum to the within-skill estimate in Panel A. Significance levels: *** $p < 0.01$, ** $p < 0.05$.

A A Theoretical Framework for AI-Driven Task Reallocation

This appendix develops a simple, stylized theoretical framework to guide the empirical analysis and illustrate how generative AI may reshape firms' outsourcing decisions, alter contract structure, induce worker reallocation, and generate concentration of market opportunities. The objective is not to deliver a fully structural account of organizational choice, but to make transparent how the technological change maps into the estimable objects in our data.

A.1 Setup

A.1.1 Tasks

A representative firm must complete a unit measure of tasks indexed by $i \in [0, 1]$ to produce output. Tasks are characterized by two attributes:

- **Automation exposure** $a_i \in [0, 1]$: the degree to which generative AI can autonomously perform task i, reducing the need for human labor input.
- **Complexity** $b_i \in [0, 1]$: the degree to which task i requires coordination across inputs, client-specific customization, iterative feedback, or exercise of judgment.

We assume that exposure and complexity are negatively correlated in the task distribution:

$$\mathbb{E}[b_i \mid a_i] = \bar{b} - \rho a_i, \quad \rho > 0. \tag{A-1}$$

This structure reflects the empirical regularity that highly automatable tasks—routine coding, translation, standardized content generation—tend to be modular and codifiable, while tasks that resist automation—strategic advising, system integration, creative direction—tend to require tacit knowledge and adaptive judgment.[27]

Let $F(a, b)$ denote the joint distribution of tasks over exposure and complexity.

A.1.2 Production Modes

Each task can be completed through one of three modes:

1. **Automation**: performed using generative AI at unit cost $\kappa_i(\theta)$.

[27] This negative correlation is consistent with the task-based literature distinguishing routine from non-routine work (Autor et al., 2003; Acemoglu and Autor, 2011). In our setting, automation exposure captures susceptibility to generative AI specifically, while complexity captures the coordination and customization requirements that make tasks difficult to fully specify in arm's-length contracts.

2. **Outsourcing**: performed by an external contractor at unit cost c_{ik} depending on contractor location k.

3. **Internal production**: performed in-house at reservation cost $\bar{w}$.

We focus on the automation–outsourcing margin, treating internal production as an outside option that absorbs tasks for which neither automation nor outsourcing is cost-effective. This focus is appropriate for our empirical setting, which observes realized outsourcing transactions on a platform marketplace.

A.1.3 Automation Technology

If task i is performed using generative AI, the unit cost is

$$\kappa_i(\theta) = \bar{\kappa} - \theta a_i, \tag{A-2}$$

where $\theta \geq 0$ indexes the capability of frontier AI systems. The parameter $\bar{\kappa}$ captures baseline costs of deploying AI (infrastructure, integration, prompt engineering, quality assurance), and θa_i captures capability-driven cost reductions that are increasing in task exposure.

Improvements in AI capability correspond to increases in θ. This formulation implies that capability improvements reduce automation costs for all tasks but do so most strongly for tasks with high exposure a_i. The discrete releases of GPT-3.5, GPT-4, and GPT-4o in our sample period can be interpreted as increases in θ that shift the automation cost frontier.

A.1.4 Outsourcing Technology

Outsourced tasks are performed by contractors located in either developing economies (f) or developed economies (d). The unit cost of outsourcing task i to destination $k \in \{f, d\}$ is

$$c_{ik} = \frac{w_k}{\phi_k(b_i)}, \tag{A-3}$$

where w_k is the destination wage and $\phi_k(b_i)$ is effective productivity, which depends on task complexity.

Wages. We assume $w_f < w_d$: developing-country contractors have lower wages, reflecting differences in outside options, cost of living, and labor market conditions. This wage differential is the fundamental driver of outsourcing to developing countries and is well documented in the literature on offshore services trade.

Productivity. We assume that productivity is similar across destinations for routine, well-specified tasks, but that developed-country contractors have a productivity advantage in complex tasks requiring judgment, coordination, or client-specific adaptation:

$$\phi_f(b) = \phi_0, \qquad \phi_d(b) = \phi_0 + \mu b, \tag{A-4}$$

where $\phi_0 > 0$ is baseline productivity common across destinations and $\mu > 0$ captures the developed-country productivity premium in complex tasks.

This structure reflects that routine, codifiable tasks such as data entry, basic translation, and standardized coding can be performed comparably well by workers in either location. The task requirements can be fully specified and transmitted digitally, and execution does not depend on physical proximity, specialized institutional knowledge, or intensive client interaction. Complex tasks, in contrast, benefit from the specialized skills, established client relationships, and institutional context more prevalent among developed-country workers.

Comparative advantage. Under these assumptions, the relative cost of developing- versus developed-country outsourcing is

$$\frac{c_{if}}{c_{id}} = \frac{w_f}{w_d} \cdot \frac{\phi_0 + \mu b_i}{\phi_0}. \tag{A-5}$$

For low-complexity tasks (b_i small), this ratio is approximately $w_f/w_d < 1$: developing countries are cheaper because the wage advantage dominates and there is no offsetting productivity differential. For high-complexity tasks (b_i large), the ratio increases and may exceed one if the developed-country productivity premium is sufficiently large relative to the wage gap.

Developing countries thus have comparative advantage in low-complexity tasks, where the wage differential determines relative costs. Developed countries have comparative advantage in high-complexity tasks, where productivity differences can offset higher wages. Combined with the negative correlation between exposure and complexity in (A-1), this implies that high-exposure tasks are disproportionately outsourced to developing countries, a pattern documented in the data.

A.1.5 Contractors

Within each destination k, contractors are heterogeneous in ability η, distributed according to $G_k(\eta)$ with support $[\underline{\eta}, \bar{\eta}]$. Ability affects productivity: a contractor with ability η performing task i in destination k has effective productivity $\eta \cdot \phi_k(b_i)$.

We assume ability and complexity are complementary:

$$\frac{\partial^2(\eta \cdot \phi_k)}{\partial \eta \, \partial b} > 0. \tag{A-6}$$

High-ability contractors have comparative advantage in complex tasks that reward skill, judgment, and adaptability. Low-ability contractors are relatively more competitive in routine tasks where the scope for differentiation is limited.

A.1.6 Contract Structure

Outsourced tasks are bundled into contracts. A contract specifies a set of tasks $S \subseteq [0, 1]$ to be performed by a single contractor. Contracting is costly: each contract incurs a fixed coordination cost $F > 0$ that captures search, matching, contract specification, monitoring, and payment processing.

The total cost of a contract covering task set S assigned to a contractor in destination k is

$$C(S; k) = F + \int_{i \in S} c_{ik}(i) \, di. \tag{A-7}$$

The fixed cost F creates an incentive to bundle tasks into contracts rather than procuring each task separately. However, the optimal scope of bundling depends on the characteristics of the tasks being outsourced.

Routine tasks can be precisely specified, easily verified, and executed independently, favoring separate procurement with specialized contractors. Complex tasks, in contrast, favor consolidation: they are interdependent (requiring coordination), incompletely specified (requiring adaptation), and benefit from accumulated client-specific knowledge. As outsourced work shifts toward complex tasks, optimal contract structure shifts toward fewer, broader engagements.

A.2 Buyer Task Allocation and Consolidation

A.2.1 Mode Choice

The firm assigns each task to its least-cost production mode. For task i, the firm compares:

$$\min \left\{ \kappa_i(\theta), \; \min_{k \in \{f,d\}} c_{ik}, \; \bar{w} \right\}. \tag{A-8}$$

Define the *automation threshold* $\hat{a}(\theta)$ as the exposure level at which automation and the

best outsourcing option have equal cost:

$$\bar{\kappa} - \theta\hat{a} = \min_k c_k(\mathbb{E}[b \mid a = \hat{a}]). \tag{A-9}$$

Tasks with exposure $a_i > \hat{a}(\theta)$ are automated; tasks with $a_i < \hat{a}(\theta)$ are outsourced (provided outsourcing is cheaper than internal production) or produced internally. The threshold $\hat{a}(\theta)$ is the margin between automation and outsourced labor in the task space.

A.2.2 Comparative Statics

Proposition 1 (Effect of AI capability on task allocation)**.** *An increase in AI capability θ has the following effects:*

1. *The automation threshold falls:* $\dfrac{\partial \hat{a}}{\partial \theta} < 0$.
2. *The measure of outsourced tasks contracts:* $\dfrac{\partial |\mathcal{I}_O|}{\partial \theta} < 0$.
3. *The average exposure of outsourced tasks falls:* $\dfrac{\partial}{\partial \theta}\mathbb{E}[a_i \mid i \in \mathcal{I}_O] < 0$.
4. *The average complexity of outsourced tasks rises:* $\dfrac{\partial}{\partial \theta}\mathbb{E}[b_i \mid i \in \mathcal{I}_O] > 0$.

Proof. Part (1): From (A-9), an increase in θ raises the left-hand side (reduces automation cost) for any fixed $\hat{a}$. To restore equality, $\hat{a}$ must fall, which raises $\bar{\kappa} - \theta\hat{a}$ back toward the outsourcing cost. Hence $\partial\hat{a}/\partial\theta < 0$.

Part (2): The outsourced set is $\mathcal{I}_O(\theta) = \{i : a_i < \hat{a}(\theta) \text{ and } \min_k c_{ik} < \bar{w}\}$. Since $\hat{a}$ falls with θ, the set of tasks satisfying $a_i < \hat{a}(\theta)$ shrinks. Hence $|\mathcal{I}_O|$ contracts.

Parts (3) and (4): The tasks removed from $\mathcal{I}_O$ as θ increases are those with exposure just below the old threshold—i.e., the highest-exposure tasks in the outsourced set. By the negative correlation in (A-1), these are also the lowest-complexity tasks on average. Removing high-exposure, low-complexity tasks from $\mathcal{I}_O$ lowers the mean exposure and raises the mean complexity of the remaining outsourced tasks. ∎

A.2.3 Destination-Specific Effects

Proposition 2 (Asymmetric displacement across destinations)**.** *Displacement of outsourcing is concentrated in developing countries. Formally, let $\mathcal{I}_f(\theta)$ and $\mathcal{I}_d(\theta)$ denote the sets of tasks outsourced to developing and developed countries, respectively. Then:*

$$\left|\frac{\partial |\mathcal{I}_f|}{\partial \theta}\right| > \left|\frac{\partial |\mathcal{I}_d|}{\partial \theta}\right|.$$

Proof. By equation (A-1), high-exposure tasks have low complexity on average. By equation (A-5), developing countries have lower costs for low-complexity tasks because their wage advantage dominates when the developed-country productivity premium μb is small. Thus, prior to the increase in θ, high-exposure tasks are disproportionately outsourced to developing countries: $\mathcal{I}_f$ is concentrated in the high-a, low-b region of the task space.

The tasks displaced by automation as θ increases are those with exposure near the threshold $\hat{a}(\theta)$, i.e., the highest-exposure tasks remaining in the outsourced set. These tasks are disproportionately drawn from $\mathcal{I}_f$. Hence, the contraction in outsourcing falls more heavily on developing-country destinations. Developed-country outsourcing $\mathcal{I}_d$ is concentrated in low-exposure, high-complexity tasks where the productivity premium justifies higher wages. These tasks remain well below the automation threshold as θ increases, so $|\mathcal{I}_d|$ is relatively unaffected. ■

The key mechanism is straightforward: developing countries' wage advantage gives them comparative advantage in routine tasks; routine tasks (high exposure, low complexity) are automated first as AI capability improves; therefore, developing countries bear the brunt of displacement.

A.2.4 Contract Consolidation

Proposition 3 (Consolidation)**.** *As AI capability increases, buyers consolidate remaining work into fewer, larger contracts.*

Proof. By Proposition 1, outsourced tasks shift toward higher complexity. Complex tasks are interdependent, harder to specify, and benefit from accumulated client knowledge, all of which favor relational contracting with a single contractor over separate procurement. Combined with fixed costs per contract, this implies consolidation. Higher complexity also implies higher contract values and more detailed specifications. ■

A.3 Worker Adjustment and Skilling

The preceding analysis characterizes firm-side responses to improvements in AI capability. We now turn to the supply side: how do workers adjust their skill portfolios in response to the changing composition of outsourced demand?

A.3.1 Skills and Skill Value

Workers supply skills to the outsourcing market. Let $\mathcal{I}$ denote the set of skills available on the platform, where each skill $i \in \mathcal{I}$ corresponds to a subset of tasks $\mathcal{T}_i \subset [0, 1]$. Skills may

overlap: a task may require multiple skills, and a worker may supply multiple skills.

Let $V_i(\theta)$ denote the expected value to a contractor of supplying skill i when AI capability is θ. This reduced-form object summarizes the equilibrium return to skill i, incorporating expected demand for the skill, expected compensation conditional on matching, and any continuation or reputation value. We do not model the determination of V_i from primitives but instead characterize how it responds to changes in θ through the demand-side mechanisms already established.

Lemma 1 (Skill values respond to AI capability)**.** *Improvements in AI capability affect skill values through two channels:*

1. ***Automation pressure**: Skills tied to high-exposure tasks experience declining demand and hence falling value as those tasks shift from outsourcing to automation.*

2. ***Reorganization complementarity**: Skills that facilitate coordination, integration, client interaction, or project management experience rising value as buyers consolidate into complex, multi-dimensional projects.*

Skills in the second category, which we term *AI-complementary*, gain value as the reorganization of outsourced work toward complex projects raises the demand for skills that support such projects. This is a composition effect operating through the demand side, distinct from direct productivity augmentation, which we explore in the extensions A.7.

A.3.2 Reskilling Technology

Workers can reallocate across skills by paying switching costs. A worker initially supplying skill j who switches to skill k incurs cost

$$c_{ji} = c_0 + \tau \cdot d(j, i), \tag{A-10}$$

where $c_0 > 0$ is a fixed component reflecting the baseline cost of acquiring any new skill, $\tau > 0$ governs the importance of skill distance, and $d(j, i) \geq 0$ measures the technological distance between skills j and k.

Distance is symmetric ($d(j, i) = d(i, j)$) and satisfies the triangle inequality. We interpret $d(j, i)$ as reflecting the transferability of human capital: skills that share common competencies, tools, or domain knowledge have low distance, while skills requiring fundamentally different expertise have high distance. In the empirical analysis, we proxy for distance using pre-AI patterns of skill co-occurrence among workers and in project requirements.

A.3.3 Skill Choice

Workers choose which skill to supply to maximize expected utility. A worker currently supplying skill j evaluates each potential destination skill k according to

$$U_{ji} = V_i(\theta) - c_{ji} + \varepsilon_{ji}, \tag{A-11}$$

where ε_{ji} is an idiosyncratic taste shock. Assume $\{\varepsilon_{ji}\}_{i\in\mathcal{I}}$ are i.i.d. Type-I extreme value with scale parameter normalized to one.

Let M_j denote the mass of workers initially supplying skill j. Under the distributional assumption, the probability that a worker from skill j transitions to skill k is

$$P_{ji} = \frac{\exp(V_i\theta) - c_{ji})}{\sum_{\ell\in\mathcal{I}} \exp(V_\ell(\theta) - c_{j\ell})}, \tag{A-12}$$

and the expected flow of workers from j to k is $\omega_{ji} = M_j \cdot P_{ji}$.

A.3.4 Gravity Structure

Taking logs of the flow equation and expanding:

$$\log \omega_{ji} = \log M_j + V_i(\theta) - c_0 - \tau d(j,i) - \log \Phi_j(\theta), \tag{A-13}$$

where $\Phi_j(\theta) = \sum_{\ell\in\mathcal{I}} \exp(V_\ell(\theta) - c_{j\ell})$ is the inclusive value from origin skill j.

With origin fixed effects α_j absorbing $\log M_j$, c_0, and $\log \Phi_j(\theta)$, we obtain the gravity specification:

$$\log \omega_{ji} = \alpha_j + V_i(\theta) - \tau \cdot d(j,i) + \nu_{ji}, \tag{A-14}$$

where ν_{ji} is a residual with conditional mean zero.

This structure is analogous to gravity equations in trade, with origin fixed effects, destination attractiveness (V_i), and bilateral frictions ($\tau \cdot d(j,i)$). It yields a directly estimable specification linking worker flows to skill proximity and destination characteristics.

Proposition 4 (Directed reskilling)**.** *Worker transitions across skills satisfy two properties:*

1. ***Proximity****: Conditional on destination value, flows are decreasing in skill distance:* $\partial \log \omega_{ji}/\partial d(j,i) = -\tau < 0$.

2. ***Direction****: Conditional on distance, flows are increasing in destination skill value:* $\partial \log \omega_{ji}/\partial V_i = 1 > 0$.

As AI capability increases, flows are directed toward skills whose value rises, particularly AI-complementary skills that benefit from contract consolidation and task reorganization.

The proximity property implies that reskilling is localized: workers move along nearby skill dimensions rather than making large jumps across unrelated domains. The direction property implies that, conditional on feasibility, workers move toward skills offering higher returns. These properties predict concentrated flows toward AI-complementary skills that are proximate to displaced skills.

Importantly, directed reskilling toward high-value destinations does not imply that workers flee automation exposure per se. Workers may move toward nominally high-exposure skill categories if those skills involve tasks that complement rather than compete with AI, such as deploying, customizing, supervising, or integrating AI systems. What matters for the direction of flows is the post-AI value $V_i(\theta)$, not the exposure level alone.

A.3.5 Heterogeneous Adaptation

Workers differ in their capacity to adjust. Let worker ability η affect switching costs:

$$c_{ji}(\eta) = c_0 + \tau \cdot d(j,i) - \gamma \cdot \eta, \tag{A-15}$$

where $\gamma > 0$ implies that high-ability workers face lower switching costs. This could reflect greater learning capacity, more transferable general skills, better access to information about emerging opportunities, or financial resources to invest in retraining.

Corollary 1 (Heterogeneous adaptation)**.** *High-ability workers are more likely to reskill in response to changes in skill values induced by AI. Among workers initially supplying skill j, the probability of transitioning to a high-value destination k is increasing in ability η.*

Proof. The transition probability is $P_{ji}(\eta) \propto \exp(V_i - c_{ji}(\eta)) = \exp(V_i - c_0 - \tau d(j,i) + \gamma\eta)$. This is increasing in η for any destination k, and the increase is proportionally larger for high-V_i destinations. Hence high-ability workers are both more likely to switch and more likely to switch to high-value destinations conditional on switching. ∎

A.4 Concentration

The above framework further implies that AI-driven reorganization increases concentration of jobs and earnings among contractors.

Proposition 5 (Increasing Concentration)**.** *An increase in AI capability raises concentration through three channels:*

1. ***Selection**: Demand contraction induces exit among contractors in high-exposure tasks, who are disproportionately low-ability by (A-6).*

2. ***Threshold effects**: Consolidation raises capability requirements; buyers bundling complex projects prefer contractors who can deliver across multiple dimensions, rationing out lower-ability contractors.*

3. ***Differential adaptation**: High-ability contractors reskill more effectively toward AI-complementary skills (Corollary 1), capturing disproportionate shares of growing categories.*

Together, these forces shift the distribution of jobs and earnings toward high-ability contractors, increasing concentration at the top of the distribution.

A.4.1 Decomposing Concentration

Changes in aggregate concentration can be decomposed into cross-skill and within-skill components. Let s_i denote skill i's share of platform earnings, and let C_i denote concentration (e.g., the top-5 share) within skill i. Aggregate concentration is $\bar{C} = \sum_i s_i C_i$, and its change decomposes as:

$$\Delta \bar{C} = \underbrace{\sum_i \bar{s}_i \Delta C_i}_{\text{within-skill}} + \underbrace{\sum_i \bar{C}_i \Delta s_i}_{\text{cross-skill}} + \text{interaction}, \tag{A-16}$$

where bars denote baseline (pre-GPT) values. The within-skill component captures changes in the distribution of earnings among contractors supplying the same skill. The cross-skill component captures shifts in platform activity across skill categories, for instance, if activity reallocates toward skills that were already highly concentrated.

The framework predicts that within-skill concentration rises in high-exposure skills through the selection and threshold mechanisms described above. The cross-skill component is ambiguous *a priori*: if high-exposure skills contract and these skills happened to be more concentrated than average, the cross-skill component would be negative (activity shifts away from concentrated skills); the reverse would generate a positive cross-skill contribution.

The within-skill component can be further decomposed to distinguish the role of worker exit from rising inequality among continuing workers. Let C_i^{pre} denote concentration in skill k computed only among workers who were active before AI introduction. Then:

$$\Delta C_i = \underbrace{(C_i - C_i^{\text{pre}})}_{\text{participation}} + \underbrace{(C_i^{\text{pre}} - \bar{C}_i)}_{\text{within-active}}. \tag{A-17}$$

The *participation* margin captures how worker entry and exit affect measured concentration. The *within-active* margin captures changes in concentration among workers who were active before AI and remain active afterward. This component isolates the pure selection and threshold effects operating within a fixed set of workers: as capability requirements rise and select workers reskill, top performers may capture a growing share, even without any change in the composition of the workforce.

The framework thus generates three testable predictions regarding concentration:

1. *Within-skill concentration rises*: High-exposure skills experience greater increases in concentration than low-exposure skills.

2. *Cross-skill reallocation is ambiguous*: The sign depends on whether high-exposure skills were more or less concentrated than average prior to AI introduction.

3. *Within-active concentration rises*: Among workers who remain active, top performers capture a growing share.

The participation margin provides a diagnostic for the nature of selection: a positive coefficient indicates that exit is concentrated among low performers, while a negative coefficient indicates that some top performers also exit, potential due to heterogeneity in adjustment costs or outside options not captured in the framework.

A.5 Extensions

We consider two extensions that address additional empirical patterns and potential alternative mechanisms.

A.5.1 AI-Assisted Production and Quality

The baseline model treats automation as full substitution: tasks are either performed by AI or by human contractors, not both. In practice, generative AI may also augment human productivity, enabling workers to complete tasks faster or produce higher-quality output.

To capture this possibility, let output quality for a contractor with ability η performing task i be:

$$q_i(\eta; \theta) = \eta \cdot \phi(b_i) \cdot \big[1 + \delta(\theta) \cdot a_i\big], \tag{A-18}$$

where $\delta(\theta) > 0$ captures AI-assisted quality improvements. The augmentation term $\delta(\theta) \cdot a_i$ is increasing in both AI capability and task exposure: AI tools provide greater assistance for tasks where AI is more capable.

This extension generates additional predictions:

Quality improvements among continuing workers. Among contractors who remain active on the platform, AI assistance raises output quality. The magnitude of improvement is larger for (i) high-exposure tasks, where AI tools are most effective, and (ii) lower-ability workers, for whom the proportional gain from assistance is greatest. A contractor with low η benefits more from the multiplicative augmentation than a contractor who was already producing high quality.

Compression of quality differentials. Prior to AI, quality differences across contractors were driven primarily by ability differences: $q_i(\eta) = \eta \cdot \phi(b_i)$. With AI assistance, the gap between high- and low-ability contractors narrows for high-exposure tasks:

$$\frac{\partial}{\partial \theta}\left[\frac{q_i(\eta_H;\theta)}{q_i(\eta_L;\theta)}\right] < 0 \quad \text{for } a_i \text{ large.} \tag{A-19}$$

This "leveling" effect is consistent with evidence that generative AI compresses productivity distributions by disproportionately benefiting lower performers (Brynjolfsson et al., 2025; Noy and Zhang, 2023).

Interaction with displacement. The augmentation and substitution effects operate simultaneously but on different margins. Substitution removes high-exposure tasks from the outsourced set entirely. Augmentation raises quality for high-exposure tasks that remain outsourced (e.g., because full automation is not yet cost-effective or because human oversight is required). The net effect on any given task depends on whether it falls above or below the automation threshold $\hat{a}(\theta)$.

A.5.2 Task Expansion

The baseline model holds the set of tasks fixed and analyzes how AI reallocates existing tasks across production modes. An alternative possibility is that AI creates new tasks or expands the scope of feasible projects.

To capture this, let the measure of tasks depend on AI capability:

$$|\mathcal{T}(\theta)| = \bar{T} + \zeta(\theta), \tag{A-20}$$

where $\bar{T}$ is the baseline task measure and $\zeta(\theta) > 0$ is an expansion term increasing in θ. New tasks created by AI—such as prompt engineering, AI system integration, model fine-tuning, AI-assisted data analysis, and human-AI workflow design—are indexed by high values of i and have characteristics that differ systematically from pre-existing tasks.

Characteristics of new tasks. We assume that AI-created tasks have moderate-to-high automation exposure (they involve working with AI systems) but also high complexity (they require judgment, customization, and coordination):

$$\mathbb{E}[a_i \mid i \in \mathcal{T}(\theta) \setminus \mathcal{T}(0)] = a^{\text{new}} > \bar{a}, \qquad \mathbb{E}[b_i \mid i \in \mathcal{T}(\theta) \setminus \mathcal{T}(0)] = b^{\text{new}} > \bar{b}. \tag{A-21}$$

These tasks are "AI-complementary" in the sense that they require human expertise to deploy, customize, and oversee AI systems rather than tasks that AI performs autonomously.

Implications. Task expansion partially offsets displacement: while AI automates some existing high-exposure tasks, it also creates new tasks that require human labor. The net effect on total outsourcing volume is ambiguous and depends on the relative magnitudes of $\partial\hat{a}/\partial\theta$ (displacement) and $\partial\zeta/\partial\theta$ (expansion).

However, the composition of outsourced work shifts unambiguously. New AI-complementary tasks are high-complexity by assumption, reinforcing the shift toward complex work documented in Proposition 1. Workers who can supply skills relevant to AI-complementary tasks, such as integration, customization, oversight, and prompt engineering, benefit from task expansion, while workers specialized in routine tasks that are automated experience displacement without offsetting demand from new task creation.

Task expansion also affects the geography of outsourcing. If AI-complementary tasks have high complexity (b^{new} large), they favor developed-country workers who have comparative advantage in complex work. This would further concentrate displacement in developing countries: not only do developing countries lose demand for routine tasks (baseline model), but the new tasks created by AI flow disproportionately to developed countries.

Empirical implications. Task expansion predicts: (i) growth in explicitly AI-related project categories (prompt engineering, AI integration, chatbot development), (ii) these projects command premium compensation reflecting high complexity, and (iii) demand for AI-related skills concentrates in developed countries or among high-ability contractors in developing countries who can meet the capability requirements. These predictions are testable using variation in AI-related keywords in project descriptions and skill tags.[28]

[28]It is worth noting that while the scope of the framework is appropriate for our setting, several margins that are unobservable in our data lie outside the model. First, wage adjustment: displacement would put downward pressure on developing-country wages, partially offsetting comparative advantage shifts. Second, reduced profitability in high-exposure skills would induce exit, while increased profitability in AI-complementary skills would attract entry. Third, the model abstracts from platform competition and the possibility that AI-driven reorganization could induce reallocation across platforms, not just within them.

A.6 Summary of Predictions

The framework generates four sets of predictions that structure the empirical analysis:

1. **Task displacement**: Outsourcing contracts in tasks with high automation exposure as AI capability improves. The automation threshold shifts inward, removing high-exposure tasks from the outsourced set and reducing overall outsourcing volume.

2. **Destination asymmetry**: Displacement is concentrated in developing-country outsourcing. Developing countries' wage advantage gives them comparative advantage in routine, low-complexity tasks, precisely the tasks with highest automation exposure. As these tasks are automated, developing-country contractors bear the brunt of displacement while developed-country outsourcing, concentrated in complex tasks, is relatively unaffected.

3. **Contract consolidation**: The composition of remaining outsourced work shifts toward complex, interdependent tasks. Buyers respond by consolidating into fewer, larger contracts that bundle related tasks under single contractors capable of coordination, adaptation, and integrated delivery. Contract values rise and project specifications become more detailed.

4. **Worker adjustment and concentration**: Workers reallocate toward technologically proximate skills with rising value, particularly AI-complementary skills that benefit from task reorganization. High-ability workers adapt more effectively, while low-ability workers face displacement and exit. The combination of selection, capability thresholds, and differential adaptation increases concentration of jobs and earnings among top performers.

These predictions provide a unified interpretation of the empirical findings. The framework rests on two main assumptions: (i) automation exposure and task complexity are negatively correlated, and (ii) developing countries have wage-driven comparative advantage in routine tasks. Both assumptions are empirically grounded and generate the qualitative patterns observed in the data without requiring strong functional form restrictions or unobserved primitives.

B Data Appendix

This appendix provides detailed information on the data sources, collection procedures, sampling design, variable construction, and limitations of the dataset. The purpose is to

ensure transparency and reproducibility and to clarify the scope and interpretation of the empirical results.

Data sources and collection. All data used in this study are collected from the API of the online labor marketplace analyzed in the paper. Data were retrieved using automated scripts that collect worker profiles, client profiles, project listings, and contract histories visible to any user browsing the platform. Collection procedures complied with the platform's rate limits and did not circumvent access controls or authentication requirements.

The dataset covers the period from January 2022 through December 2024. This window spans the introduction of generative AI tools and allows us to observe both pre-introduction outcomes and post-introduction adjustments in outsourcing activity.

The raw data consist of three linked units of observation: worker profiles, client profiles, and project-level contracts. These units are merged using platform-assigned identifiers. Prior to analysis, we perform a series of data-cleaning and consistency checks. These include removing duplicate observations arising from repeated page retrievals, verifying that project completion dates follow posting dates, standardizing currencies and timestamps, and dropping observations with missing identifiers or implausible values. These procedures ensure that the final dataset contains internally consistent and complete transaction histories.

Sampling design and cohort construction. A defining feature of online labor marketplaces is the extreme concentration of economic activity. Although the platform hosts a large number of registered users, only a small fraction ever completes a contract, and an even smaller fraction does so regularly. As a result, nearly all realized transactions are generated by a limited subset of active workers. For the purpose of studying outsourcing behavior and its response to technological change, comprehensive collection of all registered accounts is neither feasible nor informative.

To focus on the economically relevant portion of the market, we construct a fixed cohort of workers with demonstrated engagement prior to the diffusion of generative AI. For each of the platform's 2,700 skill categories, we sample up to 500 workers who completed at least one contract in that skill before November 2022, ranked by the platform's internal ranking system based on completed contracts. Over 75 percent of skill categories have fewer than 500 active workers meeting our inclusion criterion, so for these categories, typically narrow specializations, our sample includes the universe of active participants. For larger categories, our sample captures the visible portion of the market where matching actually occurs. Because workers frequently supply multiple skills (median: 8 categories), the sampling procedure in this stage yields approximately 93,000 unique individuals across all skill categories.

Since our analysis requires that we observe workers in both the pre- and post-periods,

we further refine the sample to approximately 17,000 workers who completed at least one contract on each side of November 2022. This final sample of continuously active workers accounts for approximately 95 percent of platform-wide contract volume during the sample period, consistent with the extreme concentration characteristic of digital labor markets, where a small core of participants generates the vast majority of activity (Horton, 2010; Agrawal et al., 2016).[29]

Client-worker-skill transaction data and variable definitions. For each worker in the cohort, we collect the complete history of posted and completed projects visible on the worker's profile between January 2022 and December 2024. The resulting dataset contains over 300,000 projects. Over 91 percent of these projects are cross-border transactions in which the client and worker report different countries of residence, reflecting the platform's primary function as an intermediary for international outsourcing of digitally deliverable services. Worker characteristics are drawn from publicly visible profile information and include country of residence, tenure on the platform, education and certifications when available, self-reported skills, historical job counts, and client ratings. Client characteristics are extracted from project and client pages and include client country, industry category when available, hiring history, and ratings.

Project-level information includes the project title and description, associated skill tags, contract type (hourly or fixed-price), posting and completion dates, budget ranges, number of bids, hours worked for hourly contracts, realized payments when visible, and multidimensional client evaluations covering quality, communication, timeliness, and professionalism. All monetary values are converted to a common currency using contemporaneous exchange rates. From the project-level data, we construct outcomes at multiple levels of aggregation to study different margins of adjustment. At the worker–week level, we measure the number of projects completed, total earnings, average project value, and the distribution of work across skills. At the skill–week level, we compute total project counts, total contract value, and measures of concentration across workers. At the client–week level, we aggregate the number and value of projects posted and the distribution of demand across skills. These aggregations allow us to analyze adjustment along extensive, intensive, and compositional margins of activity in the market.

Table A.1 reports the largest source countries, destination countries, and professional categories observed in the dataset and confirms that the platform is primarily used for international outsourcing of digitally deliverable services. Outsourcing flows in the data

[29] Fixing the cohort also ensures that changes in outcomes are not mechanically driven by shifts in platform traffic. To the extent that AI deters entry by workers who would have joined absent the technology, our estimates may understate aggregate displacement effects.

originate predominantly from high-income countries; the United States accounts for 27.4% of projects, followed by Australia (10%), the United Kingdom (8.1%), and Canada (5.5%). The projects are directed mostly toward developing-country destinations; approximately 60% of contracts are performed in India and Pakistan, followed by Bangladesh and Egypt. The tasks are heavily concentrated in professional services such as design, media, IT, software development, content creation, marketing, and data entry.

Table A.1: Top Source Countries, Destination Countries, and Professional Categories

Rank	Destination Country		Source Country		Professional Category	
	Name	Share (%)	Name	Share (%)	Name	Share (%)
1	India	35.35	United States	27.45	Design, Media, and Architecture	36.14
2	Pakistan	25.85	Australia	10.00	Websites, IT, and Software	34.45
3	Bangladesh	12.29	India	9.53	Writing and Content	8.63
4	Egypt	3.52	United Kingdom	8.10	Sales and Marketing	5.42
5	United States	1.68	Canada	5.48	Data Entry and Administration	5.30
	Other	21.31	Other	39.44	Other	10.07

Notes. This table reports the largest source countries, destination countries, and professional categories for cross-border digital service outsourcing observed in the platform data. Shares are computed as percentages of total completed cross-border contracts over the sample period.

C Construction of Generative AI Exposure Measure

This appendix documents the construction of our skill-level AI exposure measure. Measuring automation exposure to generative AI is inherently challenging. Existing measures differ along at least three dimensions: the unit of analysis (tasks, skills, or occupations), the underlying concept (technological capability versus realized usage), and the mapping between AI capabilities and economic activities. Because our paper studies trade in tasks mediated through an online labor platform where production is organized around narrowly defined skills rather than occupations, the choice of exposure measure is central to identification and interpretation. Our baseline analysis uses a capability-based, skill-level exposure index constructed by eliciting generative AI's ability to perform the tasks associated with each of the platform's 2,700 skill categories.

This appendix describes the elicitation procedure; the stability of exposure scores across repeated elicitations and alternative prompt specifications; comparisons with existing occupation-level indices; and the construction of AI-complementary and interactive intensity measures.

C.1 Elicitation Procedure

Prompt design. For each skill, we construct a prompt that presents the skill name and definition and then requests a quantitative assessment of generative AI's capability to per-

form the skill.[30] The full prompt template is:[31]

> `System Role: You are a generative AI scholar with deep expertise in digitally delivered services and repositories of knowledge across skill domains. Your task is to evaluate individual skills and produce indices that capture generative AI's ability to automate the skills.`
>
> `Skill: [SKILL_NAME]`
> `Broad Category: [CATEGORY]`
> `Definition: [SKILL_DEFINITION]`
>
> `On a scale from 0 to 1, how well can generative AI perform a task related to the skill [SKILL_NAME] compared to a human, with minimal human input?`
>
> `Please note: You have to perform the actual task with minimal human input. Return only the dictionary entry described earlier.`
>
> - `0 = AI cannot meaningfully perform any of these tasks`
> - `0.5 = AI can perform these tasks at roughly half the quality or completeness of a human`
> - `1 = AI can fully perform these tasks at human-level quality`
>
> `Response:`

API parameters. We query the OpenAI API using the following parameters to ensure reproducibility and minimize stochastic variation:

- **Model:** `gpt-3.5-turbo` (baseline); `gpt-4` and `gpt-4o` (robustness)
- **Temperature:** 1 (0: deterministic decoding; 2: most creative)
- **Conversation history:** None (each query is independent)

Stability across runs. To reduce stochastic variations, we repeat each prompt 100 times for each skill and average the responses. The standard error of each skill's mean exposure is approximately 0.008 at the median, an order of magnitude smaller than the cross-skill standard deviation of mean exposure scores (0.12). To put these magnitudes in context: a skill measured at 0.60 exposure has a 95 percent confidence interval of approximately

[30] Skills requiring any physical presence or offline execution (about 8% of the 2,700 skill categories) are excluded.

[31] We also considered alternative prompt variations as discussed below, and verified that exposure rankings are robust to prompt wording.

0.58–0.62. Measurement precision is therefore far finer than the cross-skill variation we exploit for identification. Any remaining measurement error is uncorrelated with true exposure and attenuates estimated effects toward zero.

C.2 Robustness to Alternative Language Models

Our baseline exposure measure is constructed using GPT-3.5, the model available at widespread commercial adoption during our sample period. To assess robustness to alternative language models, we recompute exposure scores using GPT-4 (released March 2023) and GPT-4o (released May 2024).

Table A.2 reports the comparison by reporting pairwise rank correlations across all three model versions. Exposure rankings derived from GPT-3.5 are highly correlated with those from GPT-4 and GPT-4o (0.63 and 0.56, respectively), and correlations are even stronger between GPT-4 and GPT-4o (0.75).

Figure A.1 compares skill-level exposure scores generated using GPT-3.5 to those recomputed using GPT-4. Each point corresponds to one of the platform's 2,700 skill categories, with both scores normalized to the unit interval. The figure reveals a strong positive relationship between the two measures, with a Pearson correlation of 0.61 and a Spearman rank correlation of 0.63. This close alignment indicates that the relative ranking of skills by automation exposure is largely stable across model generations.

Importantly, this consistency is not mechanical. GPT-4 and GPT-4o differ from GPT-3.5 along several dimensions, including reasoning depth, instruction following, and generalization; yet exposure rankings remain highly correlated. This suggests that our baseline measure captures the underlying technological feasibility of task automation rather than the idiosyncratic features of a particular model snapshot. Skills assessed as highly automatable by GPT-3.5 tend also to be assessed as highly automatable by more advanced models, even as overall capability expands.

At the same time, the dispersion around the 45-degree line in Figure A.1 is informative. Differences across models are most pronounced among mid-exposure skills, where automation feasibility is inherently ambiguous and more sensitive to incremental improvements in model capabilities. These are precisely the tasks for which organizational responses may unfold gradually rather than through discrete substitution. By contrast, skills at the extremes—those that are clearly automatable or clearly resistant—exhibit much tighter alignment across models.

We therefore use GPT-3.5 as our baseline model for two reasons. First, GPT-3.5 corresponds closely to the technological frontier that was widely accessible to firms at the onset

of generative AI diffusion. It underpinned the initial release of ChatGPT in November 2022 and reflects the capabilities firms and workers faced when adjustment began. Second, GPT-3.5 provides a conservative benchmark for automation feasibility. To the extent that newer models expand the set of automatable tasks, estimates based on GPT-3.5 are likely to underestimate, rather than overestimate, the longer-run scope of AI-driven reorganization. For these reasons, GPT-3.5 is well-suited for studying the early, equilibrium-shaping effects of generative AI on international outsourcing.

Table A.2: Rank Correlations of Skill-Level Exposure Scores Across Model Versions

	GPT-3.5	GPT-4	GPT-4o
GPT-3.5	1.00	0.63	0.56
GPT-4		1.00	0.75
GPT-4o			1.00

Notes. Table reports pairwise Spearman rank correlations of skill-level automation exposure scores constructed using different large language model versions. Exposure scores are computed for all platform skill categories and normalized to the unit interval.

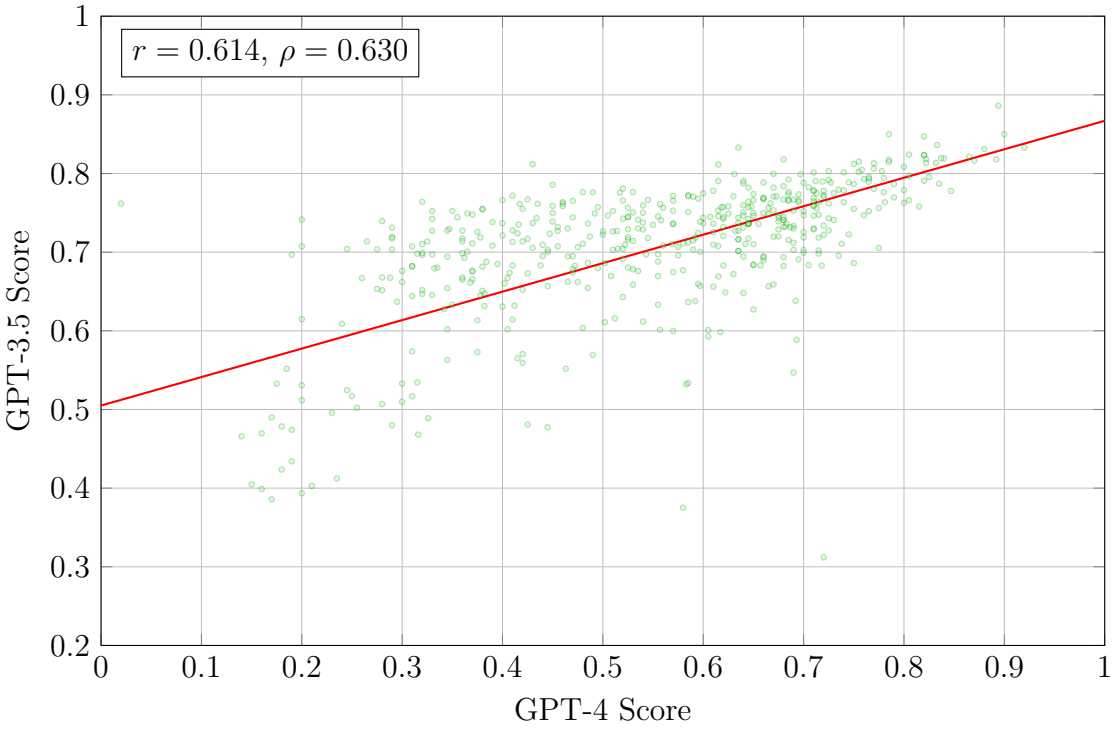


Figure A.1: Comparison of Skill-Level Exposure Scores: GPT-3.5 vs. GPT-4

Notes. Each point corresponds to a platform-defined skill category. The horizontal axis reports exposure scores constructed using GPT-4, and the vertical axis reports exposure scores constructed using GPT-3.5. Scores are normalized to the unit interval.

C.3 Validation with Internal and External Alternative Indices

Our exposure measure is designed to capture the technological scope for automation facing firms when deciding whether to outsource, automate, or reorganize tasks. A natural concern is whether our findings depend on this particular construction. To address this concern, we compare our measure with both internal alternative and leading external exposure measures developed in prior work.

First, we construct an alternative exposure measure based on AI-assisted task execution rather than full automation. This measure captures settings in which human input remains necessary, but generative AI can substantially augment productivity. The key difference from our baseline measure is that the prompt asks how well generative AI can "assist an individual in performing the tasks" instead of "perform these tasks relative to a human, with minimal human input". The two measures are highly correlated ($r = 0.87$), indicating that skills where AI can substitute for human labor largely overlap with skills where AI can augment it. This correlation suggests that the distinction between automation and augmentation is more a matter of degree than of kind, at least at the current technological frontier. We use the automation-focused measure as our baseline because it more directly captures the margin relevant to outsourcing decisions: whether firms can replace offshore labor with AI, not whether they can make that labor more productive.

In addition, we compare our skill-level measure against two prominent occupation-level indices. First, we use the exposure measures developed by Eloundou et al. (2024), who classify O*NET tasks according to whether large language models can assist with or fully automate them. Their α measure captures the share of an occupation's tasks that AI can assist with; their β measure captures tasks AI can perform with minimal human oversight. Second, we use the Anthropic AI Exposure Index, which measures realized AI usage patterns across occupations based on anonymized conversations with Claude.[32]

Each platform skill is matched to relevant O*NET occupations through a four-layer cascade. First, the skill's category (e.g., "Websites, IT & Software") is used to filter the 923 occupations down to only those in related O*NET broad groups (e.g., Computer and Mathematical + Architecture and Engineering). Next, a curated keyword dictionary checks whether the skill name contains a recognized term (E.g., "python", "accounting") and, if so, selects only occupations whose titles contain the corresponding occupation keywords within that filtered pool. If no keyword rule fires, a token-overlap method tokenises both the skill name and each candidate occupation title (after lowercasing, removing punctuation, and stripping stopwords), and accepts any occupation sharing at least one meaningful token

[32] See https://www.anthropic.com/research/economic-index.

($length > 3$). If neither the keyword dictionary nor token overlap produces any specific matches, the skill is assigned the average exposure of occupations in its domain-filtered pool.

These alternative measures differ from our baseline exposure index in important ways. They are mainly occupation-based rather than skill-based and reflect task composition within occupations rather than platform-specific skills. Nevertheless, they provide a useful benchmark for assessing whether the direction and qualitative nature of our results are robust.

Table A.3 reports correlations between our measure and the external indices. Our skill-level measure correlates positively with all three benchmarks. The Spearman correlation with Eloundou et al. (2024)'s β (automation potential) is 0.53; with their α (assistance potential), 0.67; and with the Anthropic index, 0.54. Pearson correlations show a similar pattern. These correlations indicate meaningful alignment: skills that our measure identifies as highly exposed also tend to fall within occupations that external indices classify as exposed.

At the same time, substantial residual variation remains. This divergence is expected for two reasons. First, the external measures are occupation-based, averaging across heterogeneous tasks within broad occupational categories. Our measure captures within-occupation variation that coarser indices cannot: for example, within "Software Developers," our measure distinguishes Python scripting (high exposure) from systems architecture (lower exposure). Second, the Eloundou et al. (2024) measures reflect expert judgment about task content, while the Anthropic index captures realized usage patterns, both of which may diverge from the technological capability that our measure targets. The correlations suggest our measure captures a related but distinct dimension of AI exposure.

In Appendix D.3, we re-estimate the main skill-week specifications from Section 5 using each alternative measure in place of our baseline index. Results are qualitatively similar: outsourcing volume declines significantly in high-exposure skills across all specifications. Point estimates are somewhat attenuated when using occupation-level indices, consistent with measurement error from coarser matching and within-occupation averaging. The stability of results across alternative constructions suggests our findings reflect genuine exposure-driven displacement rather than artifacts of a particular measurement approach.

Table A.3: Correlation with External Exposure Indices

External Measure	**Spearman ρ**	**Pearson r**
Eloundou et al. β (automation)	0.53	0.67
Eloundou et al. α (assistance)	0.67	0.54
Anthropic AI Exposure Index	0.54	0.52

Notes: Table reports correlations between our skill-level exposure measure and occupation-level indices from prior work.

C.4 AI Skills and Interactive Intensity

Beyond the baseline exposure measure, we construct two additional skill-level indices to capture distinct dimensions of AI-complementary task content.

AI-specific skills. We identify skills that involve working with AI systems rather than being substituted by them. These "AI-complementary" skills capture activities such as training, deploying, monitoring, or integrating AI tools into production workflows.

We classify a skill as AI-specific if its name or description contains any of the following terms or phrases:

- Explicit AI references: "artificial intelligence," "machine learning," "deep learning," "neural network," "NLP," "computer vision"
- Model operations: "model training," "model deployment," "MLOps," "model monitoring"
- Integration activities: "AI integration," "chatbot development," "prompt engineering," "LLM," "GPT"
- Data preparation: "data labeling," "annotation," "training data"

Interactive intensity. We measure the extent to which tasks within each skill require interpersonal interaction, client communication, or soft skills that are less amenable to automation. For each skill, we compute the share of associated projects (in the pre-period) whose titles or descriptions contain terms indicating interactive or relational activities.

The keyword dictionary includes:

- Communication: "meeting," "call," "discuss," "collaborate," "coordinate"
- Client interaction: "client," "stakeholder," "feedback," "revision," "consultation"
- Creativity and judgment: "creative," "brainstorm," "strategy," "advise," "recommend"
- Presentation: "present," "pitch," "demo," "workshop"

For each skill s, interactive intensity is defined as:

$$\text{Interactive}_s = \frac{\#\text{ projects in skill } s \text{ with interactive keywords}}{\#\text{ projects in skill } s} \tag{A-22}$$

computed over all projects posted in the pre-ChatGPT period (January–October 2022).

Table A.4 reports summary statistics for the three skill-level measures used in the analysis: exposure to generative AI automation, interactive intensity, and AI-related skill content. The AI exposure measure exhibits substantial dispersion across skills, with a standard deviation of 0.12 and a wide interdecile range, indicating meaningful heterogeneity in the technological feasibility of automation even within narrowly defined digital tasks. By contrast, interactive intensity is lower on average but highly skewed, reflecting that a subset of skills involves frequent client communication and coordination, while many tasks remain largely transactional. AI-specific skills remain uncommon, accounting for 5 percent of skills.

Although these measures are related conceptually, their correlations are modest. AI exposure is only weakly positively correlated with AI-related skill status, reflecting the fact that skills associated with AI span a wide range of automation potential, from routine data preparation to specialized model development and oversight. AI exposure is also positively and weakly correlated with interactive intensity. The magnitudes of these correlations are small (0.05-0.15), confirming that the three measures capture distinct dimensions of skill content rather than redundant constructions. This variation is critical for identification: it allows us to distinguish between skills that are automatable, skills that are specific to AI, and skills that are more dependent on human interaction, and to trace how generative AI reshapes outsourcing, buyer behavior, and worker adjustment along each of these margins.

Table A.4: Summary Statistics: AI Exposure Measure

Variable	**Mean**	**SD**	**P10**	**Median**	**P90**	**N**
AI Exposure	0.69	0.12	0.54	0.70	0.83	2,458
Interactive Intensity	0.23	0.25	0	0.21	0.52	2,260
AI Skill	0.05	0.22	0	0	0	2,388

Notes: Table reports summary statistics for skill-level measures. Skills with at least one completed contract in the sample period are included. AI Exposure is the mean score across 100 repeated prompts. Interactive Intensity is the share of pre-period projects containing interactive keywords. AI Skill is an indicator for skills explicitly referencing AI integration or development.

D Sensitivity Analysis

D.1 Functional Form Sensitivity: Binned Exposure

This sensitivity analysis assesses whether the aggregate contraction in outsourcing documented in the main text depends on functional-form assumptions embedded in the continuous

exposure specification. The baseline difference-in-differences model assumes that post-GPT outcomes vary linearly with exposure to generative AI. Economic theory, however, does not require such linearity: automation may operate through threshold effects, whereby tasks become abruptly unprofitable to outsource once AI capabilities exceed a certain level.

To relax this assumption, we discretize the continuous exposure score into quartiles based on its pre-GPT distribution and re-estimate the aggregate specification. The lowest-exposure quartile serves as the omitted category. This approach allows for non-linear responses and guards against the possibility that baseline results are driven by a small number of highly exposed skills. The regressions are estimated at the skill–week level using the same outcomes, fixed effects, and clustering structure as in the main analysis.

The results in Table A.5 show a clear and monotonic gradient in the post-GPT contraction of outsourcing across exposure bins. Relative to the lowest-exposure skills, higher-exposure bins experience progressively larger declines in outsourcing value, contract volume, buyer participation, and worker participation. Even moderately exposed skills (Bin 2) exhibit statistically significant declines across all outcomes, indicating that the effects of generative AI are not confined to the extreme upper tail of exposure. This exercise offers additional support to the robustness of the baseline findings documented in 5 of the main paper.

Table A.5: Aggregate Effects of Generative AI on Outsourcing to Developing Countries: Binned Exposure

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Buyers)	(4) Log(Workers)
PostGPT × Bin 2	-0.194***	-0.065**	-0.071***	-0.069***
	(0.052)	(0.028)	(0.027)	(0.025)
PostGPT × Bin 3	-0.171***	-0.090***	-0.096***	-0.098***
	(0.047)	(0.026)	(0.025)	(0.023)
PostGPT × Bin 4	-0.255***	-0.144***	-0.146***	-0.147***
	(0.045)	(0.025)	(0.024)	(0.022)
Skill FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
R^2	0.637	0.886	0.891	0.884
Observations	88,800	90,045	90,045	90,045

Notes. The dependent variables are measured at the skill-week level. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a skill-level exposure measure capturing GPT's assessed ability to substitute for workers in that skill, normalized to the unit interval. Exposure scores are divided into quartiles based on the pre-GPT distribution; Bin 1 (lowest exposure) is the omitted category. All specifications include skill fixed effects and week fixed effects. Standard errors are clustered at the skill level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Inverse Hyperbolic Sine Transformation

The main specification in Section 5 uses log-transformed dependent variables, which drops skill-week observations with zero values. This approach raises a potential concern: if AI exposure causes some skill-weeks to have zero activity, then dropping zeros may understate the true displacement effect by excluding the most severely affected observations.

Table A.6 addresses this concern using the inverse hyperbolic sine (IHS) transformation. The IHS transformation approximates the natural logarithm for large values of y but is defined at zero, allowing us to retain all observations including skill-weeks with no completed contracts. The results are quantitatively similar to the baseline log specification. The coefficients on contract volume, active buyers, and active workers are comparable in magnitude and remain highly significant. The consistency across specifications indicates that the baseline findings are not driven by selective attrition of zero-value observations. Specifically, the similar effects offer two interpretations. First, complete market exit at the skill-week level is relatively rare; most of the displacement operates on the intensive margin (reduced activity within active skill-weeks) rather than the extensive margin (skills dropping to zero). Second, to the extent that zeros do occur, they are distributed similarly across high- and low-exposure skills conditional on fixed effects, so their exclusion does not systematically bias the baseline estimates.

D.3 Benchmarking with Alternative Exposure Indices

This appendix section evaluates the robustness of our baseline results to alternative exposure measures discussed in Appendix C.3. The purpose of this exercise is twofold. First, we assess whether the aggregate contraction in cross-border outsourcing documented in the paper is sensitive to the measure of exposure to generative AI. Second, we clarify how different exposure measures correspond to distinct economic concepts (capability, task susceptibility, and occupational composition) and why the exposure measure used in the main paper is better suited to studying the reorganization of international outsourcing and skill dynamics in our setting.

For each alternative exposure measure, we re-estimate the aggregate difference-in-differences specification, Equation 1, at the skill–week level. Outcomes include the log total value of completed contracts, the log number of contracts, the log number of active buyers, and the log number of active workers in a skill subcategory. All specifications include skill fixed effects and week fixed effects, and standard errors are clustered at the skill level.

Table A.7 reports the results. Across all exposure measures, we find consistent evidence that generative AI is associated with a contraction in cross-border outsourcing to developing

Table A.6: Aggregate Effects of Outsourcing to Developing Countries

	(1) IHS(Value)	(2) IHS(Contracts)	(3) IHS(Buyers)	(4) IHS(Workers)
Post-GPT × Exposure	−1.236***	−0.523***	−0.512***	−0.496***
	(0.141)	(0.045)	(0.044)	(0.040)
Skill FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
R^2	0.685	0.869	0.871	0.865
Observations	328,536	328,536	328,536	328,536

Notes. The dependent variables are measured at the buyer-week level and transformed using the inverse hyperbolic sine (IHS). *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. *Exposure* is a skill-level exposure measure capturing GPT's assessed ability to substitute for workers in that skill, normalized to the unit interval. All specifications include skill fixed effects and week fixed effects. Standard errors, clustered at the skill level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

countries. For each measure, higher exposure predicts statistically significant declines in outsourcing value, contract volume, buyer participation, and worker participation following the introduction of generative AI. The sign and relative ordering of effects on the outcomes are remarkably stable across measures, indicating that our central qualitative conclusion—that generative AI reduces outsourcing in more automatable tasks—is not an artifact of our exposure construction.

At the same time, the magnitude of estimated effects differs across exposure measures in economically meaningful ways. The largest effects, evaluated at the mean exposure levels, are obtained using our baseline capability-based exposure score. Measures from Eloundou et al. (2024) and Anthropic yield smaller effects, though they remain statistically significant. This attenuation is expected given the coarser level of matching and aggregation: mapping occupational exposure scores or scores based on a different task taxonomy to platform skills necessarily introduces measurement error and averages cross-skill heterogeneity.

Importantly, differences in magnitude should not be interpreted as contradictions across measures. Instead, they reflect differences in what each measure captures. Occupation-based indices are constructed to assess exposure of jobs or workers in conventional labor markets and therefore average over heterogeneous tasks within occupations. This aggregation also

permits adjustment across tasks within the same occupation, which mechanically attenuates localized, within-skill effects of generative AI. Our setting, in contrast, is a marketplace for narrowly defined digital tasks, where reallocation occurs primarily within occupations and across fine-grained skills. While the sensitivity analysis shows that the aggregate contraction in outsourcing is robust to alternative exposure measures, these external indices cannot replace our baseline measure for our granular buyer and worker analyses. Our measure, by operating at the level of platform-defined skills which directly correspond to observed contracts, payments, and task bundles, is needed to trace reallocation and reskilling across the exact units of production traded in the market.

Table A.7: Benchmarking with Alternative Exposure Indices

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Buyers)	(4) Log(Workers)
Panel A. Assistance Exposure Score (This Paper)				
Post-GPT × Exposure	-0.686***	-0.473***	-0.472***	-0.456***
	(0.109)	(0.065)	(0.063)	(0.060)
Observations	88,800	90,045	90,045	90,045
R^2	0.637	0.887	0.892	0.884
Panel B. Alpha Exposure (Eloundou et al., 2024)				
Post-GPT × Exposure	-0.326**	-0.226***	-0.242***	-0.208***
	(0.082)	(0.053)	(0.052)	(0.047)
Observations	88,692	89,937	89,937	89,937
R^2	0.635	0.886	0.891	0.883
Panel C. Beta Exposure (Eloundou et al., 2024)				
Post-GPT × Exposure	-0.291**	-0.187**	-0.212***	-0.173**
	(0.135)	(0.083)	(0.081)	(0.074)
Observations	88,692	89,937	89,937	89,937
R^2	0.635	0.885	0.891	0.883
Panel D. Anthropic AI Exposure Index				
Post-GPT × Exposure	-1.013***	-0.424***	-0.421***	-0.496***
	(0.208)	(0.128)	(0.126)	(0.117)
Observations	89,098	90,352	90,352	90,352
R^2	0.634	0.885	0.890	0.882
Skill FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes

Notes. The dependent variables are measured at the skill-week level. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. Panel A uses the paper's assistance (capability-based) exposure score, constructed by eliciting GPT-3.5's ability to perform tasks associated with each skill and normalized to lie between 0 and 1. Panels B and C use the Alpha and Beta exposure measures from Eloundou et al. (2024), mapped from tasks to a subset of platform skill categories. Panel D uses the Anthropic AI Exposure Index, mapped from broader occupations to platform skill subcategories. All specifications include skill subcategory and week fixed effects. Standard errors are clustered at the skill level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.4 Buyer Heterogeneity in Task Reorganization

This subsection examines heterogeneity in buyer-side adjustment to generative AI. Specifically, the specification allows the post-GPT exposure effect to vary with two predetermined buyer characteristics measured before November 30, 2022: pre-GPT outsourcing expenditure, which proxies for the scale of a buyer's activities on the platform, and pre-GPT skill breadth, which captures the diversity of tasks previously outsourced. Because both variables are fixed in the pre-treatment period and all regressions include buyer fixed effects, identification comes from within-buyer changes over time, interacted with differential exposure to generative AI.

Consistent with Section 6, the results in Table A.8 again reveal a pattern of outsourcing reallocation. For exposed buyers at baseline levels of pre-GPT scale and scope, total outsourcing value and mean contract value increase significantly after ChatGPT's introduction, while the number of completed contracts does not. This pattern indicates that exposed buyers who remain active shift toward higher-value engagements rather than expanding outsourcing along the extensive margin, consistent with a withdrawal from routine tasks and a reorientation toward work where human input remains valuable.

Heterogeneity by pre-GPT expenditure indicates that this value upgrading is attenuated among larger buyers. The negative interaction between exposure and pre-period expenditure implies that buyers with greater pre-GPT spending experience larger declines in outsourcing contract volume and smaller increases in outsourcing value and mean contract value. One interpretation is that larger buyers have greater internal capacity to absorb exposed tasks using AI, reducing their reliance on external contractors when automation becomes feasible.

In contrast, pre-GPT skill breadth amplifies value upgrading among exposed buyers. Buyers that previously sourced a wider range of skills exhibit larger post-GPT increases in both total value and mean contract value, even as they source fewer contracts. This pattern suggests that diversity in skill requirement portfolios facilitates consolidation into fewer, higher-value engagements when generative AI reduces the profitability of routine outsourcing.

The above evidence shows that generative AI induces selective reorganization rather than uniform contraction on the demand side. Exposure to AI shifts outsourcing toward higher-value work among buyers with the organizational capacity to adapt, while buyers with greater scale are more likely to substitute away from outsourced production altogether.

D.5 Heterogeneity across Major Outsourcing Destinations

Table A.9 examines whether the aggregate contraction in outsourcing documented in the main analysis is broadly shared across major developing-country destinations or driven by a

Table A.8: Shifts in Outsourcing Demand: Heterogeneity by Buyer Scale and Scope

	(1) Log(Value)	(2) Log(Contracts)	(3) Log(Mean Value)
PostGPT × Exposure	1.466***	-0.022	1.484***
	(0.246)	(0.045)	(0.241)
PostGPT × Exposure × log(PreGPT Expenditure)	-0.404***	-0.005*	-0.397***
	(0.012)	(0.003)	(0.012)
PostGPT × Exposure × log(PreGPT Skills)	0.541***	-0.040***	0.578***
	(0.028)	(0.008)	(0.027)
Buyer FE	Yes	Yes	Yes
Source Country × week FE	Yes	Yes	Yes
R^2	0.596	0.356	0.608
Observations	92,583	94,411	92,583

Notes. The dependent variables are measured at the buyer-week level. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. log(PreGPT Expenditure) and log(PreGPT Skills) are measured over the employer's pre-GPT outsourcing history and proxy for pre-period outsourcing scale and scope. All specifications include employer fixed effects, week fixed effects, and source-country × week fixed effects. Standard errors are clustered at the employer level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

narrow subset of countries. We re-estimate the baseline skill–week difference-in-differences specification separately for the five largest destination countries on the platform—India, Pakistan, Bangladesh, Egypt, and Sri Lanka using identical fixed effects, exposure measures, and timing.

Two patterns stand out. First, the negative relationship between generative-AI exposure and outsourcing activity is present across nearly all major destinations. For India, Pakistan, Bangladesh, and Sri Lanka, higher exposure is associated with statistically significant declines in total contract value, contract volume, and active workers following the introduction of generative AI. This consistency indicates that the aggregate results are not driven by a single country or idiosyncratic institutional context, but instead reflect a common technological shock affecting digitally tradable tasks across multiple labor markets.

Second, the magnitude of the response varies systematically across countries. The contraction is largest in Bangladesh and India, countries that, prior to generative AI, specialized most heavily in routine, codifiable digital services where wage arbitrage played a central role. In contrast, effects are smaller and less precisely estimated for Egypt, where baseline outsourcing volumes are lower and skill composition is more heterogeneous. This pattern is consistent with the task-based interpretation developed in the paper: generative AI disproportionately erodes demand for tasks that formed the core of comparative advantage in certain developing-country labor markets, rather than uniformly affecting all destinations.

Importantly, the similarity of responses across value, contracts, and workers within each country suggests that these effects reflect genuine contraction in market activity rather than substitution across participants within countries. The results reinforce the interpretation that generative AI operates as a geography-unbounded form of automation, compressing cross-border arbitrage in routine digital tasks, and that its impact is felt most strongly in countries whose export profiles were most exposed to those tasks.

Table A.9: Aggregate Effects across Major Destinations

	(1) India	(2) Pakistan	(3) Bangladesh	(4) Egypt	(5) Sri Lanka
Panel A. Log(Value)					
PostGPT × Exposure	-0.657***	-0.645***	-1.112***	-0.016	-0.600*
	(0.182)	(0.210)	(0.335)	(0.277)	(0.359)
Observations	47,905	49,218	28,985	18,252	6,818
R-squared	0.585	0.555	0.506	0.382	0.452
Panel B. Log(Contracts)					
PostGPT × Exposure	-0.390***	-0.360***	-0.615***	-0.212*	-0.692***
	(0.093)	(0.130)	(0.182)	(0.123)	(0.195)
Observations	48,911	50,252	29,500	18,721	7,033
R-squared	0.869	0.853	0.810	0.667	0.723
Panel C. Log(Workers)					
PostGPT × Exposure	-0.437***	-0.364***	-0.673***	-0.240*	-0.391***
	(0.084)	(0.099)	(0.142)	(0.135)	(0.140)
Observations	48,911	50,252	29,500	18,721	7,033
R-squared	0.863	0.854	0.816	0.670	0.687
Skill FE	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes

Notes. The dependent variables are measured at the skill-week level. *Value* is the total dollar value of completed contracts, converted from local currency using contemporaneous exchange rates. *Contracts* is the total number of completed contracts. *Buyers* is the number of distinct clients posting at least one contract. *Workers* is the number of distinct individuals completing at least one contract. *Post-GPT* is an indicator equal to one for weeks after the introduction of ChatGPT on November 30, 2022. All regressions include skill and week fixed effects. Standard errors are clustered at the skill level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.6 Skill Transition Dynamics

This appendix reports a gravity-style analysis of worker skill transitions motivated by the discussion in Appendix A.3. The main paper documents that workers in high-exposure skills reallocate toward AI-complementary and interactive-intensive activities (Section 7.2). Here, we characterize the macro-structure of these transitions and test whether reallocation is shaped by adoption frictions (proxied by skill-pair relatedness) and destination-skill characteristics (automation exposure, AI-specificity, and interactive intensity).

We construct a transition matrix ω_{ij}, where each element denotes the number of workers who supplied skill i prior to November 2022 and subsequently completed at least one contract in skill j during the post-ChatGPT period (December 2022–December 2024). A worker may appear in multiple origin-destination pairs if they supplied multiple skills pre-period or transitioned to multiple destinations post-period.

Specifically, we estimate:

$$\log \omega_{ij} = \alpha_i + \beta_1 \text{Relatedness}_{ij} + \beta_2 \text{Exposure}_j + \beta_3 \text{AI}_j + \beta_4 \text{Interactive}_j + \varepsilon_{ij}, \qquad \text{(A-23)}$$

where α_i are origin-skill fixed effects, which absorb the mass of workers in each origin skill and their average outside options. The origin fixed effects imply that destination-skill coefficients (β_2 through β_4) identify pull factors conditional on a worker leaving any given origin, rather than the overall propensity to transition.

The covariates are defined as follows:

- Relatedness_{ij}: Pre-period proximity between skills i and j, measured as the log number of workers who completed contracts in both skills prior to November 2022. This proxies for switching frictions: technologically proximate skills share tooling, domain knowledge, and client networks, reducing the cost of transition.
- Exposure_j: The automation exposure of destination skill j, constructed as described in Section 3.
- AI_j: An indicator equal to one if skill j explicitly references AI-related tasks (e.g., AI integration, chatbot development, prompt engineering, model deployment).
- Interactive_j: The share of pre-period projects in skill j whose descriptions reference communication, coordination, creativity, or client-facing judgment.

Standard errors are clustered at the destination-skill level to account for correlation across origin skills transitioning to the same destination.

Table A.10 reports estimates. Column (1) includes only relatedness; columns (2)–(4) sequentially add destination characteristics; column (5) reports the full specification. Three findings emerge. First, skill-pair relatedness is a dominant predictor of transitions. The coefficient is large, positive, and significant across all specifications: an increase in relatedness raises significantly the number of transitions. Workers move primarily along technologically proximate pathways.

Second, conditional on relatedness, workers do not systematically avoid high-exposure destinations or move toward AI-specific skills. The coefficient on AI_j is near zero and statistically insignificant. This contrasts with the earning premia for AI-specific skills documented

in Section 7.3, suggesting that entry into these domains is constrained by adoption frictions, such as learning costs, tooling requirements, or buyer screening for demonstrated expertise.

Third, workers disproportionately transition toward interactive-intensive destinations. The coefficient on Interactive_j is positive and significant, indicating that conditional on technological proximity, workers reallocate toward tasks requiring communication, coordination, and client-facing judgment, activities where AI complements human input rather than substituting for it.

These findings indicate that reskilling in response to generative AI is shaped by both the feasibility of adoption and expected payoffs. Workers cannot costlessly arbitrage earnings differentials across skills; instead, they adjust primarily along lower-friction margins, reallocating toward interactive and coordination-intensive tasks that are technologically proximate to their existing expertise. The persistence of high-exposure activity and the limited movement toward high-return AI-specific skills suggest that targeted training or credentialing interventions may be necessary to facilitate broader adjustment.

Table A.10: Skill Pairwise Flows

	(1)	(2)	(3)	(4)	(5)
	Log(Flow)	Log(Flow)	Log(Flow)	Log(Flow)	Log(Flow)
Relatedness	0.828***	0.816***	0.828***	0.829***	0.817***
	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Exposure		0.590***			0.592***
		(0.089)			(0.089)
AI Skill			−0.010		−0.016
			(0.039)		(0.036)
Interactive Skill				0.157***	0.160***
				(0.058)	(0.059)
Constant	0.132***	−0.307***	0.133***	0.086***	−0.356***
	(0.007)	(0.068)	(0.008)	(0.018)	(0.069)
Source Skill FE	Yes	Yes	Yes	Yes	Yes
Observations	357,852	357,726	357,852	357,852	357,726
R^2	0.739	0.743	0.739	0.740	0.743

Notes. The dependent variable is the number of workers moving from skill i to skill j after ChatGPT, measured as the number of individuals who performed skill i pre-GPT and skill j post-GPT. *Relatedness* of skills i and j is measured by the number of individuals performing both skills before ChatGPT. *Automation exposure* measures the skill's exposure to ChatGPT, normalized to 0–1. AI skill is an indicator for AI-specific skills such as machine learning, AI integration, and AI chatbot development. *Interactive skill* is the share of projects in a skill area that require creativity, communication, and interactions. Standard errors clustered at the skill level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$