Advancing AI Capabilities and Evolving Labor Outcomes

Yong Suk Lee and Jacob Dominski
University of Notre Dame

March 20, 2025

Abstract

This study examines the evolving capabilities of AI and its impact on the labor market by examining the continuously advancing AI capabilities and real-time labor market shifts in employment status and work hours across occupations and demographic groups. Leveraging task-level data, we construct a dynamic Occupational AI Capabilities Score based on a task-level assessment using multiple state-of-the-art AI models, including OpenAI's ChatGPT 4o, Anthropic's Claude 3.5, and Meta's Llama 3. Unlike previous studies that rely on static automation risk or exposure estimates, this approach introduces a five-stage framework that evaluates how AI's capability to perform tasks in occupations changes as technology advances from traditional machine learning to agentic AI. Using this staged framework, we illustrate the extent to which occupations and employment are exposed to AI using data from the Occupational Employment and Wage Statistics. The Occupational AI Capabilities Scores are then linked to Current Population Survey data, allowing for empirical analysis of employment, unemployment, and work-hour trends. Additionally, occupation-level event studies describe how AI capabilities reshape job structures in specific professions over time. We find that increasing AI capabilities significantly increases unemployment and decreases work hours in occupation level long-differenced regressions. In individual level regressions comparing occupations highly exposed to AI to the rest before and after the release of ChatGPT we find similar results on unemployment and work hours. We find stronger effects among the highly educated and the younger workers. Overall, the results suggest that AI-driven labor shifts are occurring both at the extensive margin (unemployment) and the intensive margin (work hours), but with varied effects across different occupations and demographic groups. Ongoing monitoring of AI's labor market impact will be essential. Given the speed of AI advancements, this paper highlights the importance of real-time tracking and dynamic modeling of AI capabilities in the labor market and to anticipate labor market changes and adapt workforce policies accordingly.

Keywords: artificial intelligence, dynamic AI capabilities, prompt engineering, occupational AI capabilities, occupations, unemployment, work hours.
JEL Codes: E24, J21, J23, O33

1. Introduction

The rapid advancement of Generative AI in recent years has intensified concerns about job displacement. Initially, these concerns centered on predictive tasks where Machine Learning excelled. However, with the rise of Large Language Models (LLMs), attention has shifted to occupations involving editing, writing, coding, and content creation. More recently, AI models have begun contributing to creative work in the arts and innovation-driven activities such as research and development. Agentic AI, the most recent advancements at the time of this writing, can autonomously perform actions like searching the web or making reservations. What is particularly striking is the pace of these advancements. New capabilities are continuously being introduced in just a few months. AI's rapid progress is outpacing how organizations and workers are adapting to the technology. Many firms are unsure how best to integrate these technologies into their operations, workforce strategies, and organizational structures. This gap between AI development and organizational adaptation suggests that AI's long-term impact on labor can evolve in ways we have not yet predicted. More critically, this lag complicates the design of policy responses to mitigate AI's disruptions in the workforce and effective reskilling and training policies. Exacerbating the issue is the fact that much of the available labor market data today quickly becomes outdated. AI remains a nascent technology, and measuring its adoption and real-world use is challenging. Current efforts to collect data on AI implementation often lag behind its rapid development, and by the time data is collected and research is published, it no longer reflects the latest AI capabilities.

We seek to bridge this gap by examining occupational AI exposure that evolves with time and examine its impact on the labor market using the Current Population Survey (CPS). Our framework incorporates a five-stage model of AI capabilities, which allows for a more nuanced

and forward-looking assessment of how different AI technologies impact various occupations. In this paper, we introduce a multi-stage exposure framework that evaluates how AI exposure changes as technology advances. Each occupation is assessed across five AI capability levels:

Stage 1 – Traditional Machine Learning: Covers structured-data tasks like fraud detection, logistics optimization, and basic rule-based decision-making.

Stage 2 – Early LLMs: Includes tasks that require text-based generative AI, such as writing summaries, answering questions, or reviewing documents.

Stage 3 – Multi-Modal AI: Introduces vision, audio, and cross-modal reasoning capabilities, enabling AI to process images, charts, and videos in addition to text.

Stage 4 – Advanced Reasoning AI: Encompasses AI systems with stronger reasoning, problem-solving, and domain expertise, allowing for complex analysis and decision-making.

Stage 5 –Autonomous AI Agents: Represents AI that can operate independently, execute workflows, and complete end-to-end tasks with minimal human oversight.

This staged framework provides a dynamic trajectory of AI adoption. For example, a customer service representative's task exposure at Stage 2 may be 40%, reflecting that LLMs can handle scripted inquiries. However, by Stage 4, exposure may rise to 80%, as AI systems gain deeper conversational abilities and domain knowledge. This progression implies that AI's impact on customer service representatives' impact will likely unfold overtime as technological capabilities expand.

In addition to constructing the Occupational AI Capabilities Score, we present empirical evidence of AI's effects on employment and work hours in real-time. Specifically, we (1) analyze employment and unemployment trends of occupations based on progressing AI capabilities using the Current Population Survey (CPS); (2) examine changes in work hours and

earnings to assess whether AI is reducing job intensity rather than outright displacing workers; and (3) investigate heterogeneity in AI's effects across demographic groups, industries, and job roles—highlighting potential inequities in AI-driven job disruptions. This paper will be accompanied by a public-facing dashboard to track AI capabilities levels and labor market changes over time (this is part of the larger project and not discussed in this paper).

Recent economics research on AI and labor using randomized controlled experiments demonstrate how AI tools can influence productivity in specific occupation categories or an organization. Several of these studies find that AI has "leveling" effects, by boosting productivity among lower-skill workers (Brynjolfsson et al., 2023; Dell'Aqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Cui et al., 2024). At the same time, some recent studies have found complementarity between domain and AI expertise (Tambe 2025; Toner-Rodgers 2024). Yet, evidence of displacement effects or broader macro-labor implications has been more elusive. Scholars have not identified clear, economy-wide patterns of job loss attributable solely to AI, and the most frequently used labor market data sources do not capture AI adoption in real time. This project aims to fill some of this gap.

We take several research design approaches. First is prompt engineering to estimate the percentage of each task within an occupation that can be effectively performed by AI systems at different stages of development. Second is traditional econometrics done by linking our Occupational AI Capabilities Score to near-real-time labor market data to measure AI's effects on employment and work hours. By integrating these approaches, we aim to foster timelier, data-driven conversations about how AI reshapes the labor market and inequality, and how best to adapt policies, education, and workforce strategies to these ongoing shifts.

This project contributes to several strands of literature. First, it adds to the expanding body of research examining the impact of AI on the workforce. An expanding literature investigates how new technologies—such as information technology, robotics, and AI—affect productivity, labor demand, and organizational structures (Autor et al., 1998; Athey & Stern, 2002; Bresnahan et al., 2002; Autor et al., 2003; Bloom et al., 2014; Autor, 2015; Webb, 2020; Acemoglu & Restrepo, 2022; Agrawal et al., 2019; Acemoglu et al., 2022b; Kogan et al., 2023; Autor et al., 2024). While much of this literature relies on data lagging 5 to 10 years—often rendering AI technologies outdated—I focus on real-time labor market conditions alongside the ongoing development of new generative AI capabilities. Prior research has made significant progress in developing indirect measures of AI's impact on specific tasks and predicting which occupations may be exposed or substituted by AI (Frey & Osborne, 2017; Brynjolfsson et al., 2018; Felten et al., 2018; Webb, 2020; Eloundou et al., 2023). For example, Frey & Osborne (2017) ask machine learning scientists to predict automation probabilities of occupations based on job characteristics, Brynjolfsson et al. (2018) use occupational task descriptions and expert assessments to construct a machine-learning suitability score for various jobs. Webb (2020) leverages patent descriptions to classify occupations based on their likelihood of AI substitution, while Felten et al. (2018) develop an AI occupational impact measure using technological advances and their applicability to workplace tasks. While these studies provide valuable insights, they offer static assessments—estimating the likelihood of automation at a single point in time rather than capturing how exposure evolves as AI progresses. More recently, Eloundou et al. (2024) examined how occupational tasks are exposed to Generative AI capabilites, though, model capabilities have advanced rapidly even since then. Our study complements this literature by utilizing multiple state-of-the-art (SOTA) generative AI models to construct a dynamic Occupational AI

Capabilities Score, cross-validating across different models. Furthermore, to the best of our knowledge, this project is the first to develop a dynamic AI exposure score that evolves over time based on the changing capabilities of AI.

As with previous technological revolutions, AI is expected to reshape productivity and employment. However, historical evidence suggests that the benefits of transformative General Purpose Technologies—such as electricity and information technology—did not materialize immediately but required complementary organizational changes and investments (David, 1990; Bresnahan et al., 2002). Workforce adaptation is a critical aspect of this transformation, as AI adoption alters occupational structures and shifts demand for specific skills. Agrawal et al. (2018) conceptualize AI as a predictive tool that enhances human judgment, with its impact on an occupation depending on the extent to which AI automates core tasks or augments human expertise. The extent to which occupations are exposed to AI capabilities does not directly suggest whether those tasks or occupations will be displaced or augmented. How high AI capabilities and exposure will depend upon many factors including organizational and institutional factors, i.e., the system-wide change accompanying AI adoption (Agrawal et al. 2024). The dynamic and near real-time nature of our research allows for a more precise mapping and understanding of the evolving relationship between AI and human labor, including complementarity and substitutability across occupations, tasks, and skills.

This project also contributes by providing empirical evidence on how AI reconfigures occupations at the skill level. In particular, it speaks with the growing body of research that uses job-posting data to track technology adoption and changes in occupational and skill demand (e.g., Goldfarb et al., 2020; Acemoglu et al., 2020; Ben-Nur et al., 2022; Lee et al., 2024). (*Note: the results from the job-postings data is not presented in this version as it is still ongoing*).

Additionally, I extend this literature by analyzing more nuanced labor market dimensions, such as work hours and part-time employment (Lee et al. 2025), and by providing evidence on AI's unequal impact across different occupations and regions in the U.S.

The paper proceeds as follows. The next section describes the prompt engineering of the multi-stage Occupational AI Capabilities Score using state-of-the-art (SOTA) Generative AI models, Section 3 discusses the econometric framework to analysis the impact of AI on labor market outcomes, Section 4 provides descriptive analysis of the Occupational AI Capabilities Score and AI's potential to disrupt occupation and employment using the OEWS data, Section 5 presents results from the regression analysis analyzing the CPS data in a long difference regression, Section 6 examines individual level Difference-in-Difference regression, and Section 7 explores some event studies by occupations. Section 8 concludes by discussing our findings and implications.

2. Dynamic Occupational AI Capabilities Score

2.1 Multi-stage AI Exposures

To evaluate the exposure of each occupation to state-of-the-art (SOTA) AI at different stages, we construct an Occupational AI Capabilities Score based on task-level analysis. This score estimates the percentage of tasks within each occupation that could potentially be performed by AI models. Similar to approaches in Brynjolfsson et al. (2018) and Webb (2020), we use O*NET data, which details occupations and the tasks associated with each. We then deploy prompt engineering with multiple SOTA Generative AI models (e.g., OpenAI's ChatGPT 4o, Anthropic's Claude 3.5, Meta's Llama 3) to assess the percentage of each task that current AI systems can perform. However, in contrast to earlier studies – which typically produce a single

exposure or automation score per occupation – our approach introduces a *dynamic, multi-stage analysis*, where each stage represents a distinct level of AI capability – from predictive machine learning tools up to hypothetical fully autonomous AI agents. This staged framework allows us to dynamically assess exposure as AI evolves, rather than assigning a single static automation probability. The stages are defined as follows:

Stage 1 – Traditional Machine Learning (Basic AI): Simple, narrow AI algorithms that excel at structured, well-defined tasks. These include classical machine learning and rule-based systems for pattern recognition or data analysis. Stage 1 systems can automate routine tasks with clear input-output mappings (e.g. basic data entry, simple predictions), but lack contextual understanding or complex reasoning. These are pre-Chat GPT capabilities.

Stage 2 – Early LLMs (Advanced Narrow AI): More capable AI such as early large language models or domain-specific AI systems. Stage 2 AI can handle unstructured data to an extent – for example, generating or summarizing text and providing basic recommendations. They exhibit improved natural language understanding and pattern recognition over Stage 1. However, their reasoning is still limited, and they struggle with tasks requiring deep expertise or multi-step logic. The timing of Stage 2 coincides with the release of Chat GPT.

Stage 3 – Multi-Modal AI (Integrated Perception and Language): AI models that combine multiple modalities (text, images, audio, possibly structured data) to understand context better. A Stage 3 system, for instance, could analyze text and corresponding images or charts together. This enables handling of tasks involving visual perception or cross-modal reasoning (e.g. interpreting a graph and writing a report). Previously challenging "perception and manipulation" tasks become more tractable at this stage, as the AI can process visual inputs that traditional ML could not. Still, the AI's strategic planning and real-world decision-making

remain limited. The timing of Stage 3 corresponds to the release of multi-modal capabilities in SOTA Gen AI models.

Stage 4 – Advanced Reasoning AI (Expert-Level AI): Highly sophisticated models with strong reasoning and problem-solving abilities. Stage 4 represents AI that can perform complex cognitive tasks, approach human expert performance in specialized domains, and carry out multi-step plans. These systems can analyze ambiguous situations, consider multiple factors, and make nuanced judgments (e.g. strategic planning, complex research tasks), going beyond the pattern matching of earlier stages. Creative and social intelligence bottlenecks are partly overcome at this stage, as the AI can generate original ideas or recommendations and engage in interactive decision support. However, the AI is not fully autonomous; it may still require human oversight for judgment calls or ethical decisions. The timing of Stage 4 corresponds to the release of reasoning capabilities in SOTA Gen AI models.

Stage 5 –Autonomous AI Agents (General AI): This is the most advanced stage, involving AI agents with general-purpose capabilities and a high degree of autonomy. Stage 5 AI can execute entire workflows independently, combining perception, reasoning, and action (including physical actions via robotics or software tools). Such an agent could, for example, receive a high-level goal and carry it out by planning tasks, gathering information, executing decisions, and monitoring outcomes with minimal human intervention. At this stage, nearly all tasks that are technically feasible to automate on the computer can be handled by AI, including complex problem-solving and real-time adaptation. A Stage 5 AI corresponds to the theoretical end-point of automation for non-physical tasks, although in practice some tasks may still require a human touch for oversight, creativity or empathy. Stage 5 capabilities are currently being developed and only rudimentary versions of AI agents currently being introduced as of March 2025.

2.2 O*NET Task Evaluation

For each occupation, we examine its constituent work tasks as defined in O*NET and assess AI capabilities at each of the five stages. The O*NET database provides detailed information about occupations in the U.S. economy, including the specific tasks performed in each occupation and the relevance of each task. This dataset allows for a granular understanding of which tasks could be affected by AI. Additionally, we utilize data from the Occupational Employment and Wage Statistics (OEWS) program, which aggregates employment and wage information by occupation. This data helps analyze overall employment trends and wage fluctuations across occupations with varying levels of AI exposure.

Concretely, every task is evaluated to determine what percentage of that task could be effectively performed by an AI system at a given stage. This yields a series of five exposure estimates per task (one for each stage). At Stage 1, many complex or nuanced tasks might have 0% exposure (meaning basic ML cannot perform them at all), whereas straightforward and routine tasks might show significant exposure. As we progress to Stage 5, the exposure percentage for each task can increase – reflecting the growing portion of the task that AI can handle as capabilities improve. For example, a task like "prepare a routine budget report" might be 30% automatable with basic ML, but nearly 90% automatable with an advanced reasoning AI. In contrast, a task involving high-level strategic decision-making might remain only partially automatable even by Stage 5, if human judgment or ethics are essential.

Each task-stage combination in our analysis includes not just a percentage, but also a rationale and a confidence level. The *exposure percentage* estimates the fraction of the task's work (i.e., probability of the task being done satisfactorily) by AI at that stage. The *rationale* is a

short explanation of why the AI can or cannot perform the task at that level, detailing the specific capabilities or limitations of the AI. For instance, a Stage 2 rationale might note that "an early LLM can draft a generic email based on given points, but cannot reliably handle nuanced customer inquiries," explaining a partial exposure score. The *confidence level* (e.g. High, Medium, Low) captures the assessor's certainty in that judgment – recognizing that these evaluations involve some subjective estimation of AI capabilities. We employed a systematic prompt-engineering approach to generate these evaluations: a large language model was guided through a standardized prompt for each task, asking it to step through the five stages and output the percentage, explanation, and confidence for each stage. This ensured consistency in how criteria were applied across hundreds of tasks. The use of an LLM in the assessment process is inspired by recent studies that use AI predictions for evaluating task automation potential (Eloundou et al. 2024).

By compiling the stage-by-stage task exposures, we can construct a dynamic exposure profile for each occupation. Rather than a single automation risk number, an occupation is characterized by five automation capabilities percentages (one per stage), each being an aggregate of its constituent tasks' exposures (weighted by each task's importance, i.e., share of work in that occupation). For example, a job might be 20% exposed at Stage 1, 45% at Stage 3, and 80% at Stage 5, reflecting how initially only routine tasks are automatable, and eventually even complex tasks become mostly doable by AI. This approach captures not only how much of the job could be automated, but when and at what level of AI – providing a richer picture of AI's capabilities to perform tasks as models progress.

2.3 Models Used and System Prompts

In conducting the above prompt engineering, we query large language models including ChatGPT, Claude, and Llama to estimate the percentage of each task within an occupation that can be effectively performed by a SOTA AI systems at each point in time. The Appendix provides examples of system prompts used to create the exposure score. The model-generated task-level performance estimates are then aggregated to the occupation level, producing an overall exposure score that quantifies the proportion of tasks within an occupation that AI can perform. This results in a comprehensive Occupational Gen AI Exposure Score, which provides a quantitative measure of AI's potential impact on various jobs.

As of now, we have used OpenAI's ChatGPT-4o, Anthropic's Claude 3.5, and Meta's Llama 3.0. However, we plan to continue to cross examine using additional models (Mistral, Google, and DeepSeek's models), and assess robustness and cross-validate across different prompting strategies. An important aspect in this process is the significant variation in API costs among different models. For instance, running queries across all 20,000 or so tasks in the O*NET database costs approximately 10 times more when using a closed-source model like ChatGPT compared to an open-source model like Llama, and can every vary two to three-folds among closed-source models. we document and analyze differences in model performance and cost efficiency to assess trade-offs between accuracy, cost, and accessibility.

3. Econometric Framework

We then examine how recent advances in AI has impacted employment outcomes, including unemployment rates, work hours, and labor demand (TBD) by occupation. To conduct the empirical analysis, we merge the Occupational AI Capabilities Score with the Current Population Survey (CPS) data. The CPS, a monthly survey of approximately 60,000 U.S. households

conducted by the Bureau of Labor Statistics (BLS), collects a wide range of employment-related information. The BLS uses CPS data to produce timely reports on employment, unemployment, and other labor market trends. By leveraging the high-frequency nature of CPS data, we can examine the near real-time effects of AI on occupational employment dynamics. In addition to CPS data, we are in the process of incorporating proprietary job postings data from Lightcast (formerly Burning Glass Technologies). This dataset provides information on a near-universal set of job vacancies and details the skills in demand, allowing me to analyze the impact of AI exposure on labor and skill demand. We have the job postings data through mid-2024 and plan to purchase regular updates to ensure continued near real-time analysis.

The first set of regressions is at the occupation level and uses the Occupational AI Capabilities Score at a specific stage and runs the following regression where we average the CPS variables over a 3 month period:

$$Y_{o,Sep-Nov\,2024} - Y_{o,Sep-Nov\,2022} = \alpha + \beta AI_{o,StageN} + X_{o,Sep-Nov\,2022}\Pi + \varepsilon_{it} \qquad (1)$$

where Y_{ot} is the outcome variable of interest (e.g., dummy for unemployment or hours worked last week) for occupation *o* averaged over the time period *t*, $AI_{o,StageNt}$ is the Occupational AI Capabilities Score at model stage N (N=1,2,3). This regression estimates how exposure in a specific stage predicts the average change in our outcome variables from the three months before the introduction of ChatGPT to the same three months (Sept., Oct., Nov.) in 2024. Controls X_{ot} are the averages of each control proportion for the baseline 3 months (Sept. - Nov. 2022). We examine variants of the above model my examining different time periods and model development stages.

The second model is a first-differenced model that uses the change in the Occupational AI Capabilities Score at different stages, e.g., Stage 1 and Stage 3.

$$Y_{o,Sep-Nov\,2024} - Y_{o,Sep-Nov\,2022} = \alpha + \beta(AI_{o,StageM} - AI_{o,StageN}) + (X_{o,Sep-Nov\,2024} - X_{o,Sep-Nov\,2022})\Pi + \varepsilon_{it} \qquad (2)$$

This regression estimates the effect of the change in AI capabilities between two defined stages (e.g., Stage 1 and Stage 3) on the change in our outcome variable between ~Q4 of 2022 and ~Q4 of 2024. We also use the difference between control variables in ~Q4 of 2024 and ~Q4 of 2022 as our control variables. Additionally, we investigate the heterogeneous effects of Generative AI across different groups, including variations by education, sex, age, and region. Furthermore, we analyze how employment patterns across various occupations have evolved over time and what insights these empirical trends reveal about AI's labor market impact.

We also conduct individual level Difference-in-Difference regressions by categorize occupations based on their AI capabilities scores. Specifically, we identify occupations that fall into the top quartile of occupations in terms of AI capabilities. These occupations are most exposed to automation or augmentation, providing valuable insights into the potential displacement or transformation of jobs due to AI advancements. Then we analyze the following Difference-in-Difference regression model

$$Y_{iot} = \alpha + \beta AI_o Post_t + X_{it}\Pi + \mu_o + \eta_t + \varepsilon_{it} \qquad (3)$$

where AIo is an indicator for the top quartile of occupations with AI capabilities at different stages, and $Post_t$ is after the release of ChatGPT. We also explore dynamic treatment effects using below equation where D_t is a dummy variable indicating each month before and after ChatGPT release month.

$$Y_{iot} = \alpha + \beta_t \sum_{-12}^{20} AI_o D_t + X_{it}\Pi + \mu_o + \eta_t + \varepsilon_{it} \quad (3)$$

4. Occupational AI Capabilities Across Stages and Models

The results of our AI exposure analysis are illustrated in Figure 1, which presents the evolution of AI exposure scores across different models over time. The mean exposure score aggregates the share of each task that AI models can perform at each stage of development, revealing a clear trend of growing capabilities. As AI progresses from Stage 1 (basic machine learning) to Stage 5 (autonomous AI agents), exposure scores rise significantly across all models. Anthropic Claude 3.5 Sonnet consistently assesses the highest exposure scores, followed by OpenAI ChatGPT 4o. Meta Llama 3 70B is lower but this may be due to the different system prompt used as well as the earlier date of the assessment. The sharpest increase in exposure scores occurs between Stage 4 (Reasoning models, December 2024) and Stage 5 (Agentic AI, March 2025), which shows the rapid increase in AI's ability to perform tasks if agentic AI were to be widely adopted.

The corresponding Table 1 provides a detailed breakdown of AI exposure scores across five stages, including standard deviations, minimum and maximum values, and confidence estimates. At Stage 1, the mean exposure scores are relatively low across all models, with Claude at 30.76, ChatGPT at 25.13, and Llama at 15.70, reflecting the limited scope of traditional machine learning in automating occupational tasks. However, as AI capabilities expand, exposure scores rise dramatically. By Stage 3, which represents multi-modal AI capable of integrating language, perception, and structured data, Claude 3.5 reaches 56.74, while ChatGPT4o follows at 48.95. At State 4, AI systems can meaningfully contribute to a wider range of occupational tasks, particularly those involving cross-modal reasoning and analysis. By Stage 5, the most advanced AI models approach near-complete automation potential, with Claude predicting a mean exposure score of 84.84 and ChatGPT of 81.59. These values suggest that, under a hypothetical scenario where AI is widely adopted, a substantial proportion of occupational tasks could be performed autonomously. The models' confidence estimates further

reinforce this conclusion. At earlier stages, models are less confident in their predictions with medium-confidence ratings dominating (e.g., 89% for ChatGPT in Stage 3). However, by Stage 5, high-confidence predictions become more prevalent (71% for ChatGPT), indicating increased certainty in AI's ability to perform complex tasks reliably. Overall, the figure and table illustrate the dynamic trajectory of AI's occupational impact. As models progress through successive stages, they exhibit increasingly higher exposure scores, reflecting their growing potential to perform human tasks. Appendix Figure 1 shows the correlation between OpenAI ChatGPT 4o and Claude 3.5 Sonnet scores illustrating the strong correlation between the predictions across different models.

Figure 2 illustrates AI exposure intensity across five stages of AI development by depicting the percentage of occupations at or above a given AI exposure score. As AI capabilities progress from Stage 1 to Stage 5, the curves shift rightward, indicating that higher proportions of occupations are increasingly exposed to automation potential. The overall pattern confirms a progressive expansion in AI's ability to automate occupational tasks as technology advances.

To better understand what these exposure scores imply for the labor market we link our occupational AI exposure scores with employment and wage data from the OEWS (Occupational Employment and Wage Statistics).

Figure 3 illustrates AI exposure intensity three models by plotting the percentage of occupations (upper panel) and employment (lower panel) at or above a given AI exposure score at Stage 3. If we examine the middle blue line, over 40% of occupations and 50% of workers have an exposure score of 50%. To better examine the implications on the workforce in Figure 4 we depict employment distribution across varying levels of exposure to artificial intelligence (AI), grouped into 12 occupation categories at different model stages (Stage 1, Stage 3, and

Stage 5). The employment data, classified according to the 2018 Standard Occupational Classification (SOC), is grouped into 12 occupational categories. At Stage 1, AI exposure is spread across occupational groups at levels less than 30% exposure with some office and sales occupations being exposed at a higher rate. At Stage 3, the distribution shifts, with certain groups, particularly "Office and Administrative Support," and "Sales and Related Occupations," and now “Management” and “Computing and Engineering” occupations exhibiting increased proportions of their workforce exposed to higher AI levels, suggesting heightened vulnerability or transformation potential due to AI. By Stage 5, these trends become even more pronounced, highlighting substantial shifts in the employment distribution with white-collar occupations revealing very high exposure to AI, suggesting these groups face a greater risk of displacement or restructuring. Conversely, occupation groups like "Service Occupations" and "Construction and Extraction Occupations" display lower exposure, reflecting their comparative resilience or limited susceptibility to AI disruption. The progression through the stages shows the uneven distribution of AI exposure and how occupational groups of different employment size may experience significant disruptions soon.

Figure 5 presents binned scatter plots illustrating the relationship between AI exposure and log yearly income across five AI capability stages, based on ChatGPT 4o. Each panel corresponds to a different stage of AI development, from Stage 1 (basic machine learning) to Stage 5 (fully autonomous AI). In the earlier stages, AI exposure appears relatively low and dispersed, particularly among lower-income occupations. But at this stage exposure tends to decrease linearly with income. As AI sophistication increases, a more pronounced upward trend emerges, particularly in Stages 4 and 5, where higher-income occupations exhibit greater AI exposure. This pattern suggests that advanced AI models are increasingly capable of automating

tasks across both low- and high-income professions, potentially shifting the nature of automation impact beyond traditionally routine jobs to more complex, well-compensated roles.

Figure 6 presents scatter plots depicting the relationship between AI exposure and log total employment across five AI capability stages. In the early stages, AI exposure is relatively low and scattered, with a weak correlation to employment levels. As AI systems become more advanced, particularly in Stages 4 and 5, a clearer positive relationship emerges, suggesting that occupations with higher employment levels tend to experience greater AI exposure. This trend indicates that as AI models become more capable, they increasingly impact a broader range of highly employed occupations, reflecting their growing ability to automate tasks across industries with significant workforce participation.

4.1 Top 40 exposure occupations by stages

Table 2 presents the top 40 occupations ranked by their AI exposure scores across three different stages of AI capability, based on ChatGPT 4o's assessments. The results provide insights into which jobs are most susceptible to automation at various levels of AI exposure, revealing patterns in how AI's impact evolves over time. At Stage 1, occupations with the highest AI exposure scores primarily consist of clerical, data entry, and administrative support roles. Accounting Clerks rank highest, with an AI exposure score of 64.55, followed by Tellers (61.10), Billing Clerks (58.15), and Data Entry Keyers (56.75). These roles generally involve structured, repetitive tasks that align well with traditional rule-based machine learning systems. Many clerical positions—including Payroll Clerks, Statistical Assistants, and Order Clerks—also appear among the top-ranked occupations, reflecting the susceptibility of structured data processing tasks to automation at this level. By Stage 3, AI has advanced to integrate multiple modalities, allowing for improved contextual understanding and reasoning. The ranking remains

similar in terms of occupation types, but exposure scores increase significantly, indicating AI's greater ability to handle more complex, cross-modal tasks. Accounting Clerks continue to rank first (80.32), followed closely by Tellers (77.74), Billing Clerks (77.08), and Credit Clerks (76.42). The presence of Medical Secretaries, Technical Assistants, and Financial Clerks within the top-ranked occupations suggests that AI can now perform more nuanced text-based and decision-support tasks, including processing financial records, summarizing medical notes, and assisting with customer inquiries. At Stage 5, AI reaches its most advanced stage, exhibiting near-autonomous capability in performing occupational tasks. The exposure scores increase significantly, with many occupations exceeding 90% exposure. Accounting Clerks again top the list with a score of 96.84, followed by Tellers (96.62) and Payroll Clerks (96.58). The dominance of clerical, administrative, and data-heavy roles remains apparent, as tasks such as record-keeping, transaction processing, and report generation can now be executed autonomously with minimal human supervision. The high exposure scores suggest that these occupations could be almost entirely automated at this stage. A notable pattern in the Stage 5 rankings is the increased presence of roles that require more complex, structured decision-making. Insurance Claims Clerks, HR Assistants, and Tax Preparers all appear among the top-exposed occupations, reflecting AI's ability to handle not only routine administrative tasks but also more dynamic, judgment-based activities that involve interpreting policies, processing claims, and generating compliance reports.

The results highlight several important trends in AI's impact on occupational exposure. First, clerical and administrative roles consistently rank among the most exposed occupations across all AI stages, reinforcing the notion that structured, rule-based work is particularly susceptible to automation. Second, the exposure scores increase significantly as AI progresses,

with many occupations experiencing a shift from moderate exposure at Stage 1 to near-total automation potential by Stage 5. Finally, while early-stage AI primarily affects highly repetitive tasks, later-stage AI is capable of handling more complex decision-support functions, including those found in finance, human resources, and insurance.

These findings suggest that the automation potential of AI will not be limited to basic administrative work but will extend into higher-order professional functions as AI models become more sophisticated. The staged approach underscores the dynamic nature of occupational exposure, demonstrating how AI's impact evolves with advancements in capability.

4.2 Bottom 40 exposure occupations by stages

Table 3 presents the bottom 40 occupations ranked by their AI exposure scores at different stages, based on OpenAI's assessments. These occupations exhibit the lowest susceptibility to automation across AI's progression, highlighting roles that are less amenable to replacement by current and future AI systems. At Stage 1, the occupations with the lowest AI exposure scores predominantly involve physical, manual, and highly dexterous tasks. Ambulance drivers, actors, and drywall installers top the list, with exposure scores of 7.32, 7.30, and 7.50, respectively. Other low-exposure roles include musicians, paramedics, firefighters, and police officers, all of which involve high levels of physical interaction, real-time decision-making, and unpredictability—traits that traditional machine learning systems struggle to handle. Many of these jobs also require situational awareness and rapid adaptability to changing environments, which remain outside the capabilities of early-stage AI. By Stage 3, which represents multi-modal AI with improved contextual understanding, the exposure scores for these occupations increase slightly but remain relatively low. Drywall installers, ambulance drivers, and fence erectors remain among the lowest-ranked occupations, with scores under 20. While AI at this

stage can process more complex inputs such as visual and auditory data, these occupations continue to demand hands-on dexterity and problem-solving in dynamic environments. Roles such as housekeepers, tree trimmers, and janitors begin to see modest increases in exposure, suggesting that AI-assisted automation may play a supplementary role in certain routine aspects of these jobs. However, occupations involving specialized craftsmanship, physical coordination, and environmental adaptation—such as cement finishers, millwrights, and construction laborers—still exhibit minimal exposure to AI automation. At Stage 5, when AI reaches a level of general intelligence and autonomy, the exposure scores for these occupations rise more noticeably but remain significantly lower than those of clerical and administrative professions. Roofers, drywall installers, and telecom installers continue to rank among the least exposed occupations, with exposure scores below 30. Many of the lowest-ranked occupations involve skilled trades, such as brickmasons, ironworkers, and vehicle mechanics, which require intricate manual work and real-world problem-solving that AI struggles to replicate fully. Additionally, roles such as firefighters, paramedics, and truck mechanics maintain relatively low exposure, reinforcing the notion that emergency response, physical intervention, and real-time situational assessment remain difficult to automate completely.

Overall, the results in Table 3 indicate that occupations with the lowest AI exposure scores tend to share common characteristics: reliance on physical labor, complex motor skills, high situational variability, and human judgment in real-time environments. While AI advancements may introduce tools to assist workers in these fields, full automation remains unlikely, even in a scenario where general AI reaches an advanced stage. The staged framework highlights that while AI can augment certain aspects of these jobs, complete displacement

remains challenging due to the inherently unpredictable and hands-on nature of the work involved.

Before turning to regression results, we selected a few occupations to visually examined their employment trends in Figure 9. In particular, we examine software developers, customer service representatives, and maids to find any visual cues. Interestingly we find an increasing trend in unemployment among software developers but stable employment numbers. This is likely an artifact of CPS where occupation is reported even when one is unemployed. On the other hand we see a considerable decline in people employed as customer service representative, but unemployment rate holding steady. This potentially suggests the rise of gig economy jobs among this population. For maids we don't see a noticeable change after the pandemic, which is consistent with their works being least exposed to AI.

5. Regression results

The regression analysis presented in Tables 4 to 7 investigates the relationship between occupational AI capabilities and occupation level labor market outcomes over different periods (three-month averages from September-November and six-month averages from June-November between 2022 and 2024). Table 4 demonstrates that higher exposure to AI capabilities (ChatGPT and Claude) is significantly associated with negative employment outcomes. Specifically, without controls (Panel A), increased exposure to ChatGPT correlates with a slight but significant decline in employment ($\beta = -0.10$, $p < 0.01$), though the impact on unemployment rate and hours worked in main jobs is not statistically significant. However, when we add demographic and regional controls (Panel B), the negative effect on employment strengthens ($\beta = -0.487$, $p < 0.01$), and effects on hours worked become statistically significant and negative (β

= -0.018, $p < 0.01$). Similarly, Table 5 confirms these trends over a longer averaging period of six months, maintaining significance and similar magnitudes of negative effects on employment, hours worked, and unemployment rates, especially when demographic and regional controls are introduced. Analysis using Stage 3 capabilities in Tables 6 and 7 further substantiate these results, showing consistent and somewhat amplified negative effects associated with greater exposure to advanced AI capabilities. The impact of AI exposure remains negative, statistically significant, and slightly larger in magnitude, indicating the robustness of these relationships.

Table 8, which assesses changes in AI capabilities from Stage 1 to Stage 3, provides detailed insights into the incremental effects of advancing AI capabilities. The long difference regression results in Table 8 addresses some of the omitted variable concerns that may existed in the previous tables. Results indicate significant negative impacts on employment ($\beta = -0.073$ without controls and $\beta = -0.056$ with controls, both $p < 0.01$), as well as statistically significant reductions in hours worked in both main and all jobs. Additionally, Claude exposure presents similar negative patterns, underscoring the consistent adverse labor market impacts of advancing AI technology. The regression results consistently suggest negative impacts across employment measures, demonstrating clearly that advancements in AI capabilities are significantly associated with declining labor market outcomes, even after controlling for relevant demographic and regional factors. Overall, the regression results consistently show that progressing AI capabilities are associated with notable and consistently negative impacts on employment outcomes, with the severity of these effects becoming more pronounced as AI advances and exposure duration increases.

6. Difference-in-Differences Results

Figure 7 examines unemployment trends over time for occupations grouped into quartiles based on their ChatGPT 4o Stage 1 AI exposure scores. The x-axis represents time in months from early 2021 to early 2025, while the y-axis tracks the unemployment rate. The four lines correspond to different levels of AI exposure, with the lowest exposure quartile (blue) consistently experiencing the lowest unemployment rates, while the highest exposure quartile (black dashed line) shows relatively higher unemployment throughout the period. The gap between these groups seems to increase after the introduction of ChatGPT, suggesting a potential impact of AI exposure on unemployment. I next investigate this in a regression framework.

Table 9 presents difference-in-difference estimates examining the impact of AI exposure on labor market outcomes by comparing occupations in the top quartile of AI exposure at Stage 1 and Stage 3 before and after AI advancements. The outcome variables include unemployment and hours worked in the main job. The table controls for occupation and month-year fixed effects across all specifications, with additional demographic controls included in the latter columns. The results suggest a statistically significant increase in unemployment for occupations with high AI exposure. In columns (1) and (5), the interaction term for Post x Top quartile of AI Exposure (Stage 1) is positive and significant ($\beta = 0.00172$, $p<0.05$; $\beta = 0.00197$, $p<0.01$), indicating that occupations with high AI exposure in the early stage experienced a modest but significant rise in unemployment after AI adoption. Similarly, for occupations with high exposure at Stage 3, the Post x Top quartile of AI Exposure coefficient is also positive and significant ($\beta = 0.00182$, $p<0.05$; $\beta = 0.00195$, $p<0.01$), suggesting that as AI progressed to more advanced stages, the unemployment effects persisted.

The impact on hours worked is more mixed. The coefficients for Stage 1 AI exposure in columns (2) and (6) are positive but statistically insignificant, suggesting that early-stage AI

exposure did not meaningfully affect hours worked. However, for occupations in the top quartile of AI exposure at Stage 3, AI exposure is associated with a significant reduction in hours worked in the main job ($\beta = -0.0491$, $p<0.1$ in column 4; $\beta = -0.0737$, $p<0.1$ in column 8). This finding suggests that as AI capabilities advance, affected workers in highly exposed occupations may experience reductions in working hours, potentially reflecting job restructuring or reduced labor demand.

The inclusion of additional demographic controls (columns 5-8) does not substantially alter the key findings, though it slightly increases the explanatory power (R^2 rises from 0.014 to 0.019 for unemployment and from 0.112 to 0.133 for hours worked). The results reinforce the notion that AI exposure is associated with rising unemployment and reduced work hours, particularly in later AI development stages. These findings highlight the evolving labor market impact of AI, where initial exposure may have limited effects on work intensity, but as AI systems become more capable, displacement effects become more pronounced.

Table 10 presents difference-in-difference (DD) estimates analyzing heterogeneous effects of AI exposure on unemployment across demographic and educational subgroups. The results indicate that AI exposure affects different subgroups in varying ways. In Panel A, which examines the impact of Stage 1 AI exposure, the high school or less subgroup experiences a significant increase in unemployment ($\beta = 0.00419$, $p<0.01$), while those with a bachelor's degree or more see a decline in unemployment ($\beta = -0.00225$, $p<0.05$), suggesting that lower-educated workers are more vulnerable to early-stage AI displacement, whereas higher-educated workers may have greater adaptability. Among age groups, the unemployment effects are strongest for workers aged 30 or less ($\beta = 0.00474$, $p<0.01$), while effects for middle-aged and older workers are not statistically significant. In terms of gender, AI exposure is associated with

a significant increase in unemployment for women ($\beta = 0.00271$, $p<0.01$), while the effect for men is smaller and not significant.

In Panel B, which evaluates Stage 3 AI exposure, some patterns shift. The unemployment effect for high school or less workers remains positive but weaker ($\beta = 0.00261$, $p<0.1$), while for workers with more than a bachelor's degree, the effect becomes positive and highly significant ($\beta = 0.00796$, $p<0.01$), suggesting that more advanced AI systems begin to impact even highly educated occupations. Younger workers (≤30) continue to experience higher unemployment risk ($\beta = 0.00453$, $p<0.01$), while effects for middle-aged and older workers remain small and insignificant. Notably, both men ($\beta = 0.00213$, $p<0.05$) and women ($\beta = 0.00206$, $p<0.05$) experience significant increases in unemployment under Stage 3 exposure, indicating broader labor market disruptions as AI capabilities advance.

7. Occupation-Level Investigative Event Study Analysis

While the difference-in-differences approach provides valuable insights into the overall labor market effects of Generative AI, conducting an investigative event studies analysis offers a complementary perspective that addresses several important considerations. First, it helps mitigate some of the limitations and criticisms of using imputed measures of GenAI exposure, allowing for a more targeted occupation-level comparative analysis. Second, an event study framework can more explicitly capture the dynamic nature of AI adoption across different occupations, which may be affected to varying degrees and at different times. This is particularly relevant for occupations widely cited in popular press—such as programming, legal research, writing and editing, and marketing—that are often portrayed as being at the forefront of AI disruption. By examining these specific occupations within a dynamic treatment effect

framework, researchers can directly test anecdotal claims and better understand the temporal pattern of AI's impact, ultimately deepening and refining the insights gained from the initial difference-in-differences results.

Building on these insights, the event studies further reveal how different tiers within the same occupational category can experience varying levels of AI-related disruption. An event-study within the Computer and Mathematical Occupations classification (Figure 11) indicates that software developers and programmers face relatively higher unemployment after ChatGPT's release, compared to other computing roles (e.g., network administrators, database architects). In business-oriented fields like accounting or marketing, managerial roles show relatively smaller shifts in unemployment rates compared to the more operational or analyst positions (e.g., financial analysts, marketing specialists), hinting that automation may more immediately target tasks rather than strategic oversight (Figure 12).

Additionally, there are many mixed evidence across various occupations as can be seen in the Appendix Figures. Some occupations, such as paralegals, appear more vulnerable, while editors, graphic designers, and writers have not yet shown consistent or sustained unemployment increases—highlighting a gap between media-driven narratives and early empirical evidence. These nuances underscore the value of an occupation-focused event study approach, allowing researchers to pinpoint which types of tasks and roles are most at risk and offering a clearer picture of AI's uneven impact.

8. Conclusion

This study provides a comprehensive analysis of the labor market impact of Generative AI, incorporating both methodological advancements and empirical findings. By constructing an

Occupational AI Exposure Score through a task-level assessment of multiple state-of-the-art AI models—including OpenAI's ChatGPT, Meta's Llama, and Anthropic's Claude—this research moves beyond static automation risk estimates to introduce a dynamic, multi-stage framework. This framework categorizes AI capabilities into five distinct stages, ranging from traditional machine learning to fully autonomous AI agents, enabling a more nuanced understanding of how occupational exposure evolves as AI technology advances.

The results highlight the rapid evolution of AI capabilities and their increasing impact on occupational tasks. As AI progresses through development stages, exposure scores rise significantly, indicating a growing automation potential. The sharpest increase occurs between December 2024 and March 2025, suggesting accelerated advancements in AI's ability to perform economically relevant tasks. Among AI models, Anthropic Claude 3.5 Sonnet consistently scores highest in exposure assessments, followed by OpenAI ChatGPT 4o, while Meta Llama 3 70B scores lower, possibly due to different assessment parameters. The staged framework of AI exposure provides critical insights into how automation potential expands over time. Early-stage AI (Stage 1) primarily impacts clerical and administrative roles with structured, rule-based tasks. However, as AI advances to Stage 3, multimodal capabilities enable greater occupational exposure, particularly in analytical and decision-support roles. By Stage 5, AI models demonstrate near-complete automation potential, affecting both low- and high-income professions. The correlation between AI exposure and income varies non-linearly, with mid-to-high-income occupations experiencing the greatest exposure, contrasting with previous waves of technological adoption.

The regression analysis suggests that increased AI exposure is associated with modest rises in unemployment, particularly among middle-aged, older, and male workers. However, AI

also contributes to increased work intensity in some occupations, reflecting augmentation rather than outright job loss. While lower-educated workers face higher displacement risks in early AI stages, advanced AI models begin to impact even highly educated professionals. Difference-in-differences estimates reinforce these findings, showing a significant increase in unemployment for occupations in the highest AI exposure quartiles post-AI adoption.

These results indicate that AI-driven labor shifts are occurring both at the extensive margin (unemployment) and the intensive margin (work hours), but in varied ways across demographic and occupational groups. Notably, the analysis finds that AI exposure does not uniformly translate to job losses; rather, its impact is contingent on specific workforce characteristics, industry structures, and the stage of AI development relevant to each occupation.

The study's findings have several important implications. First, they highlight the urgency of proactive workforce policies that focus on retraining and reskilling initiatives, particularly for workers in knowledge-intensive professions who are more susceptible to AI-driven job displacement. Second, the results challenge conventional assumptions that automation primarily affects lower-skilled jobs, instead showing that Generative AI is increasingly reshaping professional, business, and analytical roles. Third, the dynamic nature of AI's impact underscores the need for continuous labor market monitoring, as technological advancements may lead to further disruptions in ways that static models fail to anticipate.

Previous research has shown that the impact of new technologies, such as electricity and information and communication technologies, on labor productivity and demand often emerges with considerable delay (David, 1990; Bresnahan et al., 2002), and AI is expected to follow a similar trajectory. Therefore, near real-time analysis of AI's impact on the labor market, combined with public dissemination through a dashboard, will provide valuable insights. These

findings will have important implications not only for government policies aimed at mitigating AI's negative effects on workers but also for initiatives focused on education and workforce upskilling in an AI-driven economy. The creation of publicly accessible tools to track AI exposure and employment shifts in real-time could provide policymakers, educators, and businesses with actionable insights to navigate AI-driven labor transformations effectively.

Future research should build upon this framework by integrating additional labor market indicators, such as wage effects and job transitions, to provide a more comprehensive picture of AI's economic impact. Moreover, as AI continues to evolve, there is a pressing need to assess the effects of newer AI models, including agentic and multimodal systems, which may further expand AI's occupational reach. Additionally, firm-level analyses could complement this study by examining how businesses adapt their workforce strategies in response to AI adoption.

As Generative AI rapidly advances, its implications for the workforce will continue to unfold. The findings of this study underscore the importance of timely, adaptive policy responses that not only mitigate displacement risks but also harness AI's potential to enhance productivity and job quality. By developing a forward-looking approach to AI exposure, this research provides a foundation for better understanding, preparing for, and shaping the future of work in the age of AI.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022b. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." Journal of Labor Economics 40

Acemoglu, Daron, and Pascual Restrepo. 2019. "The wrong kind of AI? Artificial intelligence and the Future of Labour Demand." Cambridge Journal of Regions, Economy and Society 13 (1): 25–35.

Acemoglu, Daron, and Pascual Restrepo. 2022. "Tasks, Automation, and the Rise in U.S. Wage Inequality." Econometrica.

Agrawal, A., J. S. Gans, and A. Goldfarb. 2019. "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction." Journal of Economic Perspectives 33: 31–50.

Athey, Susan, and Scott Stern. 2002. "The Impact of Information Technology on Emergency Health Care Outcomes." RAND Journal of Economics 33 (3): 399–432

Autor, David, Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" The Quarterly Journal of Economics

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." The Quarterly Journal of Economics 118 (4): 1279–1333.

Autor, David. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." Journal of Economic Perspectives 29: 3–30.

David, Paul A. 1990. "The Dynamo and The Computer: An Historical Perspective On T." *American Economic Review* 80 (2):355.

Ben-Ner, Avner, Ainhoa Urtasun, and Bledi Taska. 2023. "Effects of New Technologies on Work: The Case of Additive Manufacturing. *ILR Review*, 76(2): 255-289.

Bresnahan, Timothy F., Brynjolfsson, Erik, and Hitt, Lorin M. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-level Evidence." *Quarterly Journal of Economics* 117: 339-376.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2018. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." In Economics of Artificial Intelligence. University of Chicago Press.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2018. "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *AEA Papers and Proceedings* 108: 43-47.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2023. "Generative AI at Work."

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2024. "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers."

Dell'Aqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." HBS Working Paper 24-013. ,

Eloundou, Tyna, Sam Manning, Pamela Mishkin and Daniel Rock. 2024. "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." Science. Vol 384.

Felten, Edward W., Manav Raj, and Robert Seamans. 2018. "A Method to Link Advances in Artificial Intelligence to Occupational Abilities." *AEA Papers and Proceedings* 108: 54-57.

Felten, Edward W. and Raj, Manav and Seamans, Robert. 2019. "The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization." *Mimeo*.

Frank, Morgan R., AUtor, Dafvid, Bessen, James E., Brynjolfsson, Erik, Cebrian, Manuel, Deming, David J., Feldman, Maryann, Groh, Matthew, Lobo, Jose, Moro, Esteban, Wang, Dashun, Youn, Hyejin, and Rahwan, Lyad. 2019. "Toward understanding the impact of artificial intelligence on labor." *PNAS* 116 (14): 6531-6539.

Graetz, Georg, and Guy Michaels. 2018. Robots at work. Review of Economics and Statistics 100(5): 753–68.

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis. 2020. "Artificial Intelligence in Health Care? Evidence from Online Job Postings." *AEA Papers and Proceedings* 110: 400-404.

Humlum, Anders. 2021. Robot adoption and labor market dynamics. Mimeo

Koch, Michael, Ilya Manuylov, and Marcel Smolka. 2019. Robots and firms. CESifo Working Paper No. 7608. CESifo Network: Promoting Economic Research.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller. 2023. "Technology 43 and Labor Displacement: Evidence from Linking Patents with Worker-Level Data.

Lee, Yong Suk, Taekyun Kim, Sukwoong Choi, and Wonjoon Kim. 2022. "When Does AI Pay Off? AI-Adoption Intensity, Complementary Investments, and R&D Strategy" *Technovation*. 118: 102590.

Lee, Yong Suk, Ye Ji Kee. 2024. "Artificial Intelligence and Changes in Occupational and Skill Demand: Evidence from Banking." Mimeo

Lee, Yong Suk, Toshi Iizuka, and Karen Eggleston. 2025. "Robots and Labor in Nursing Homes." *Labour Economics*. 92.

Noy, Shakked, and Whitney Zhang. 2023. “Experimental evidence on the productivity effects of generative artificial intelligence.” Science 381 (6654): 187–192.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.

Seamans, Robert, and Raj, Manav. 2018. “Artificial Intelligence, Labor, Productivity, and the Need for Firm-Level Data.” *NBER Working Paper*: 1-14.

Tambe, Prasanna. 2025. “Workforce Reskilling for AI: How Algorithmic Bilinguals Drive Organizational Value.” Forthcoming at *Management Science.*

Toner-Rodgers, Aidan. 2024. “Artificial Intelligence, Scientific Discovery, and Product Innovation.” Working Paper.

Trajtenberg, Manuel. 2018. “AI as the Next GPT: A Political-Economy Perspective.” *NBER Working Paper Series 24245*.

Webb, Michael. 2020. “The Impact of Artificial Intelligence on the Labor Market.” *Mimeo*.

Figure 1. Occupational AI Exposure

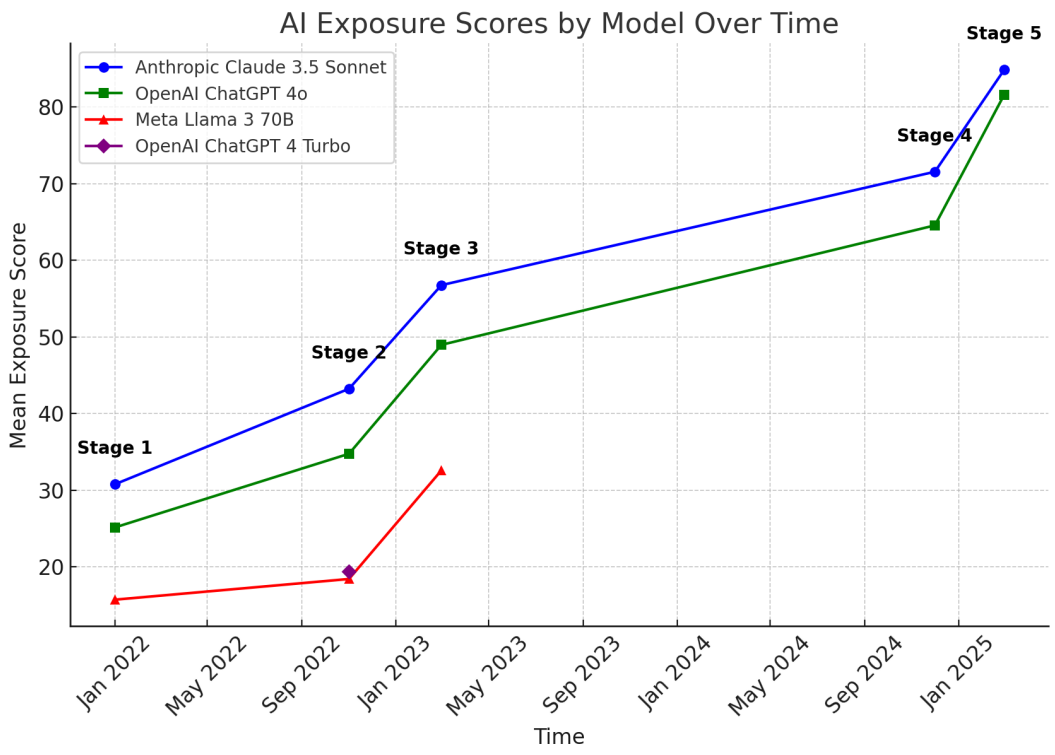


Figure 2. Cumulative AI Exposure across Stages (ChatGPT 4o)

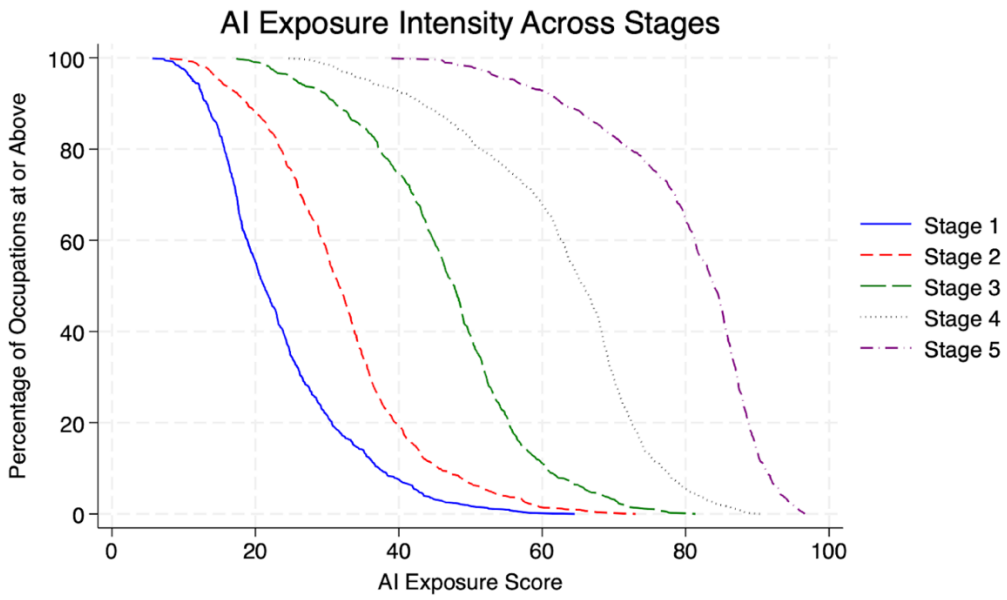


Figure 3. Cumulative AI Exposure by Model (Stage 3)

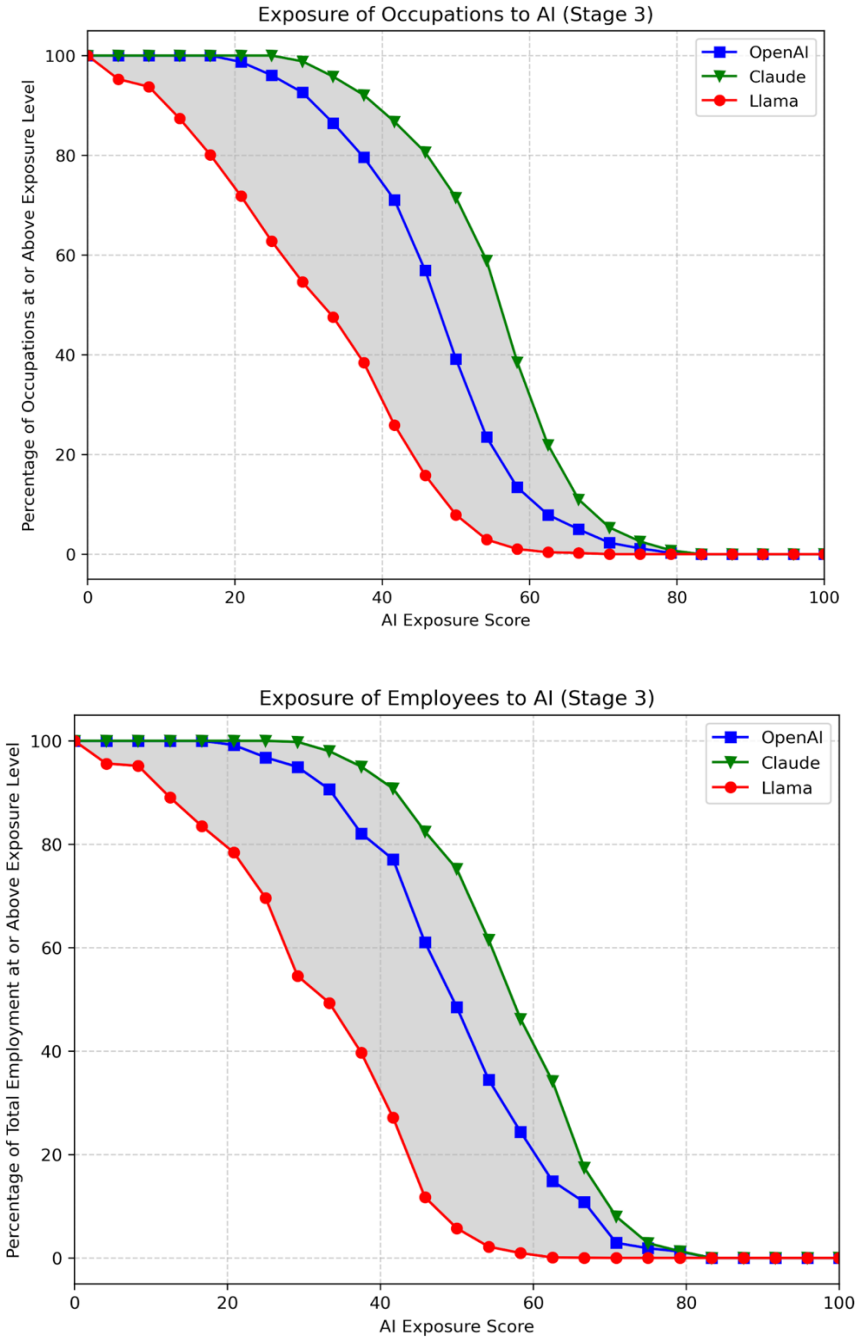


Figure 4. Employment Distribution Across AI Exposure by Stage

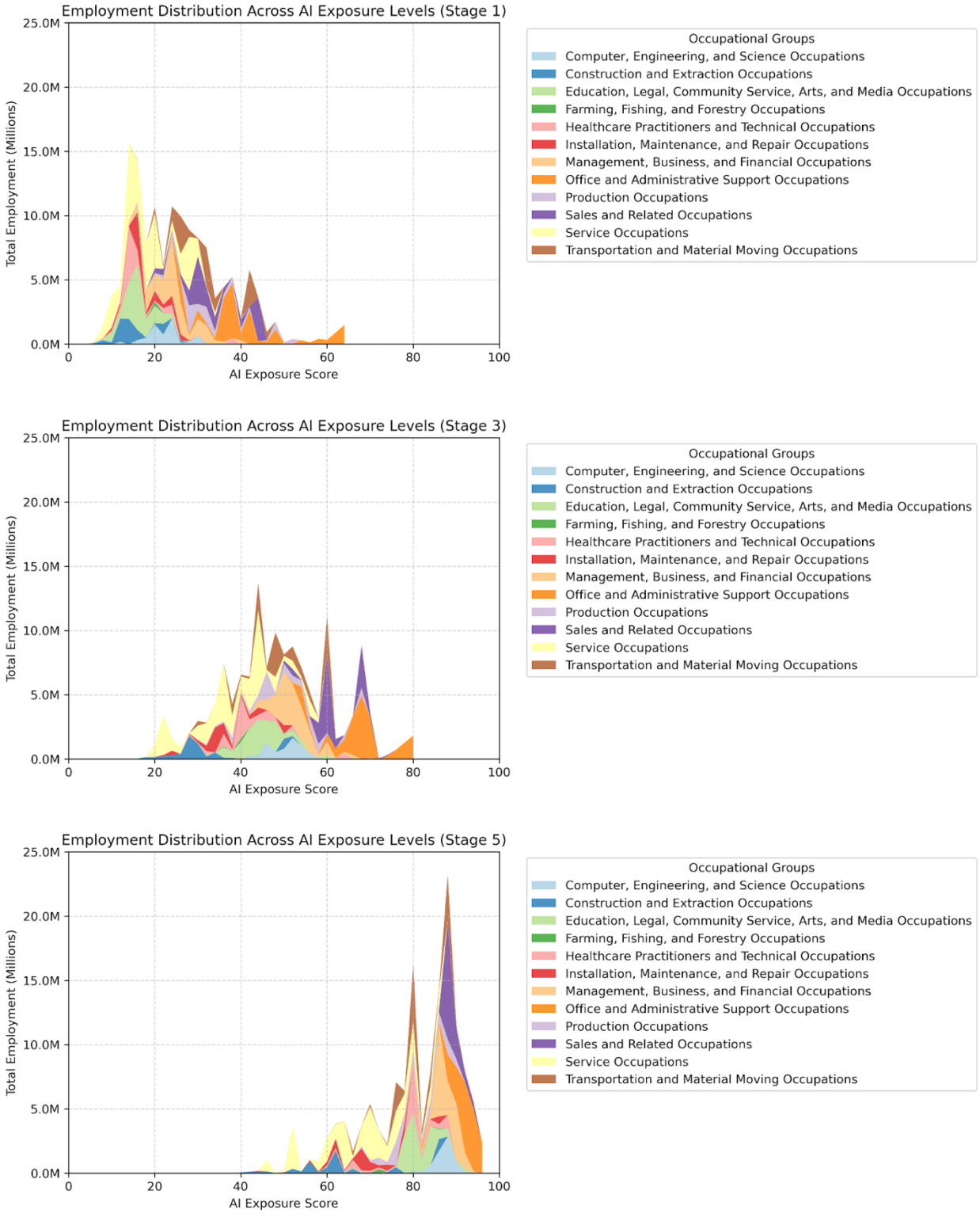


Figure 5. AI exposure by income and model stage (based on ChatGPT 4o)

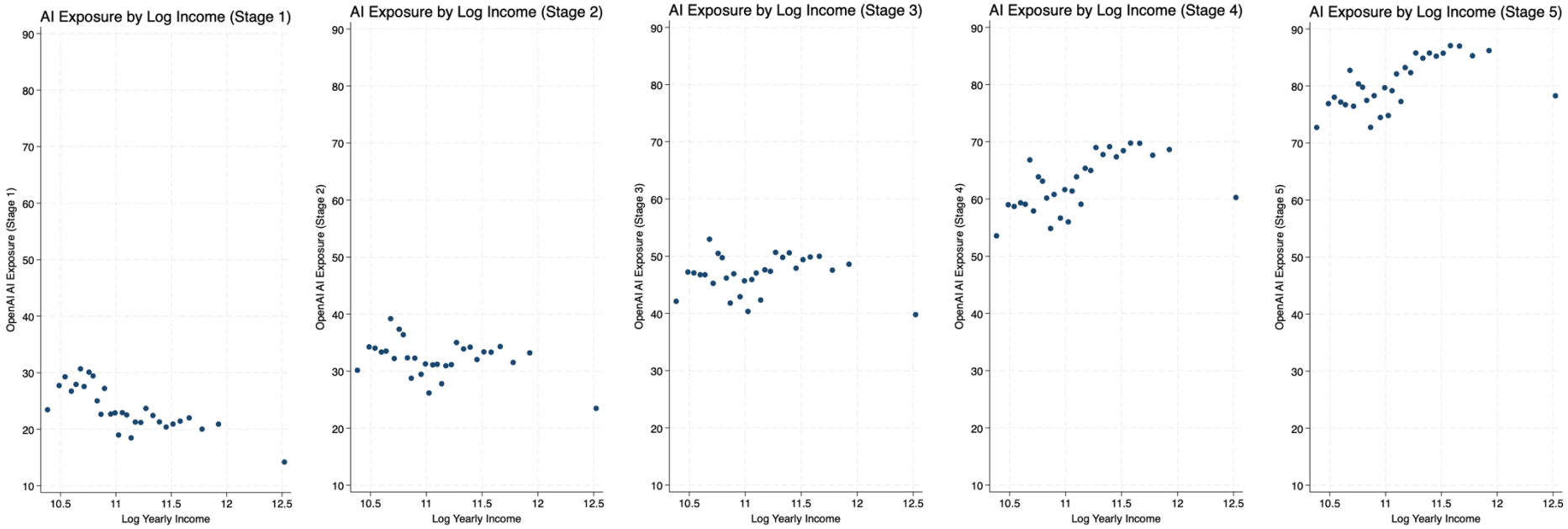


Figure 6. AI exposure by employment and model stage (based on ChatGPT 4o)

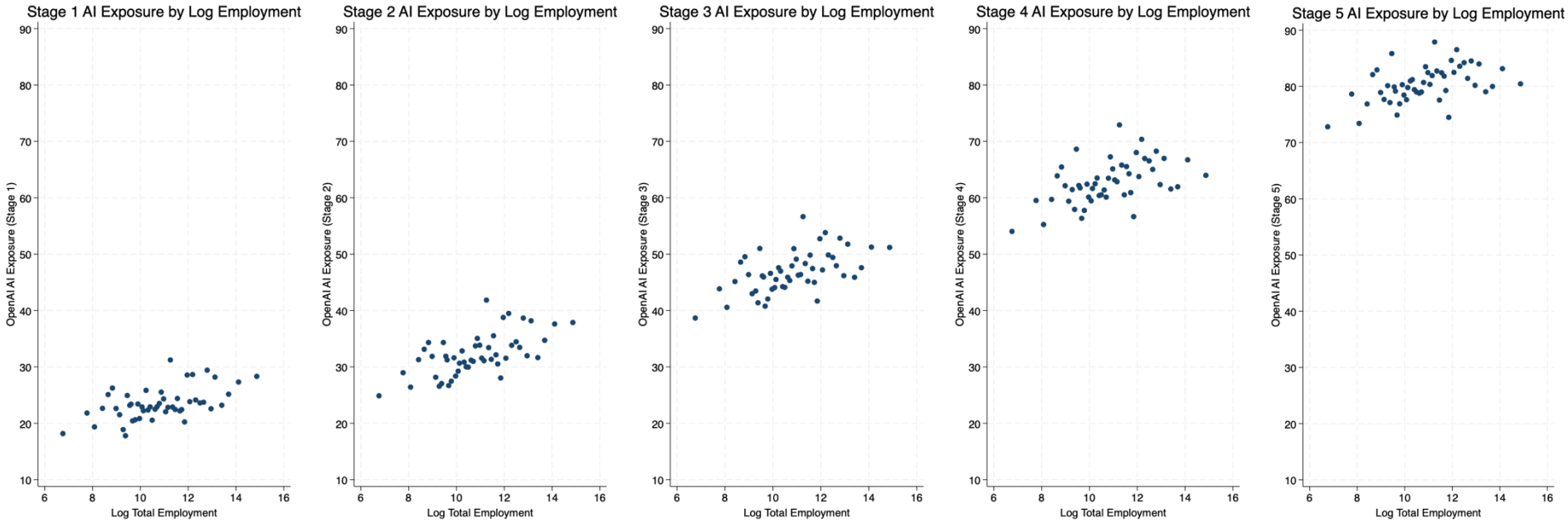


Figure 7. Unemployment Trends by Occupations grouped by exposure (Quartiles Based on ChatGPT 4o Stage 1 Exposure Scores)

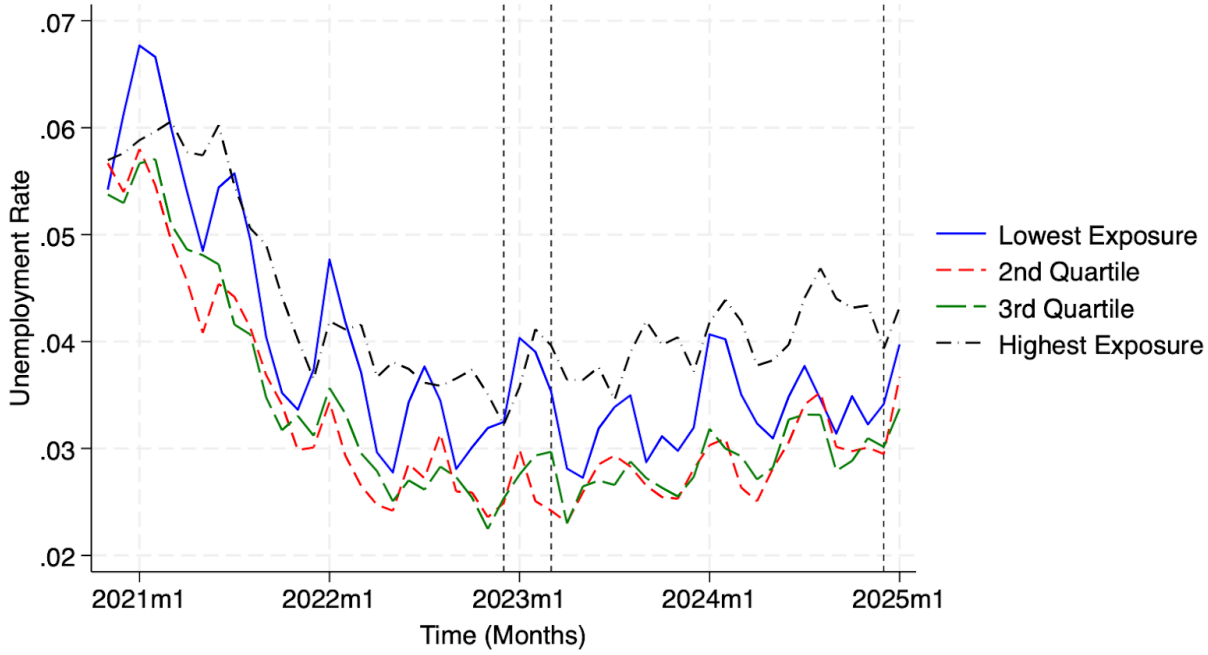
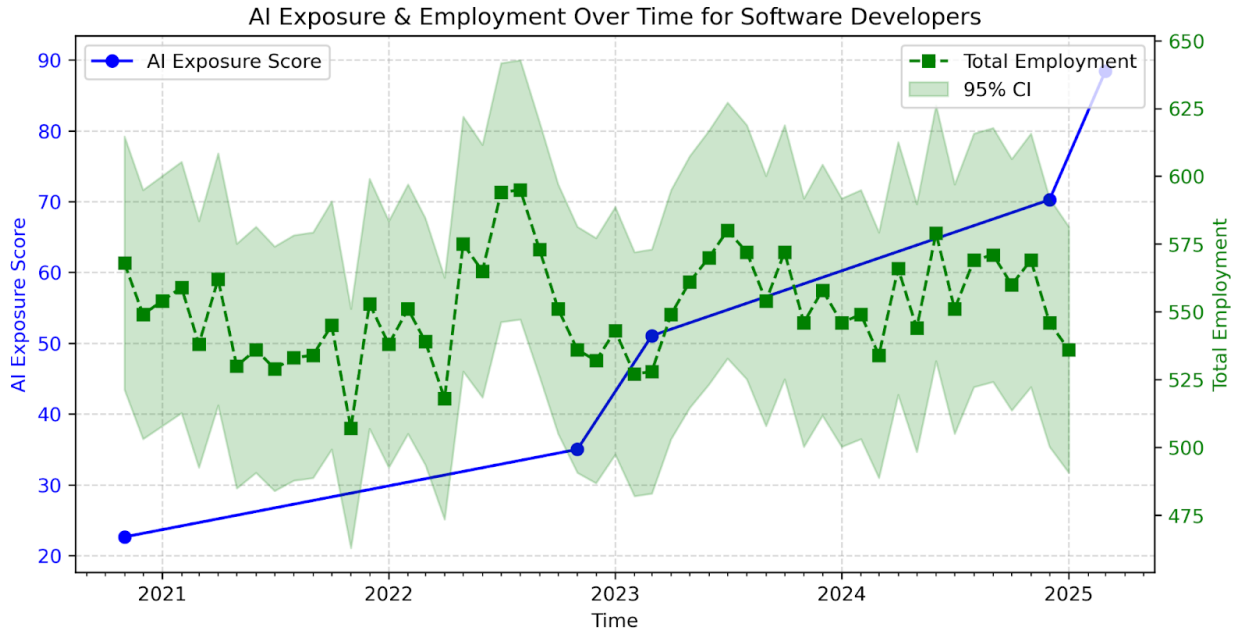
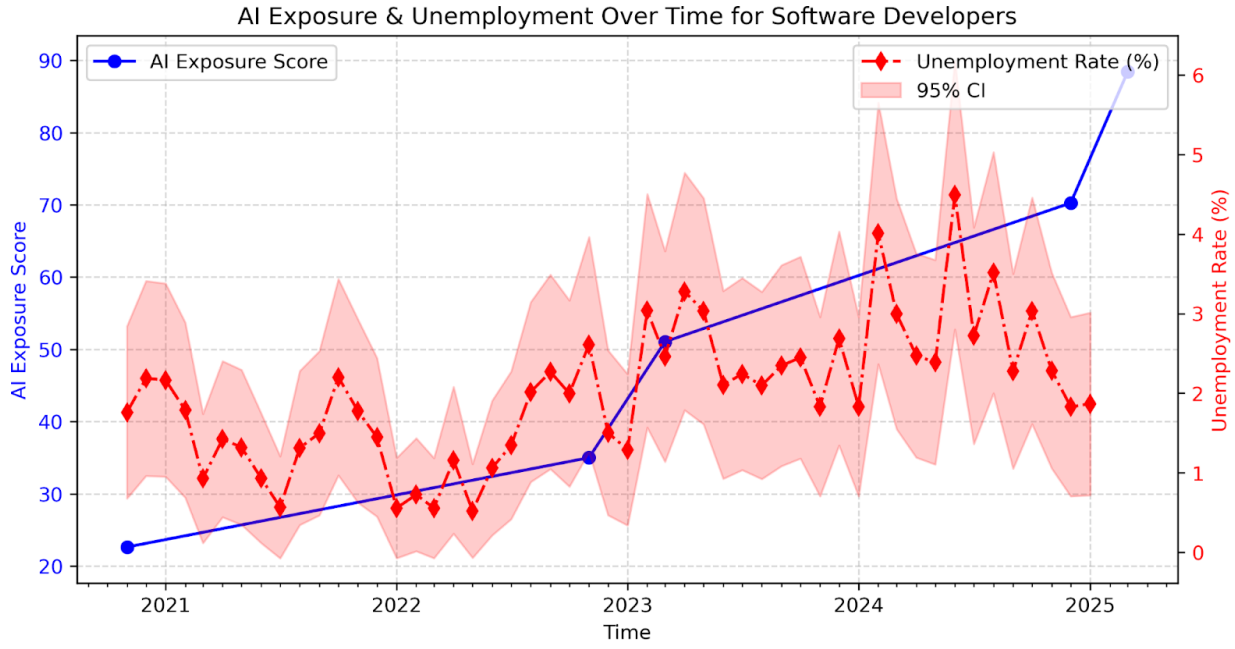


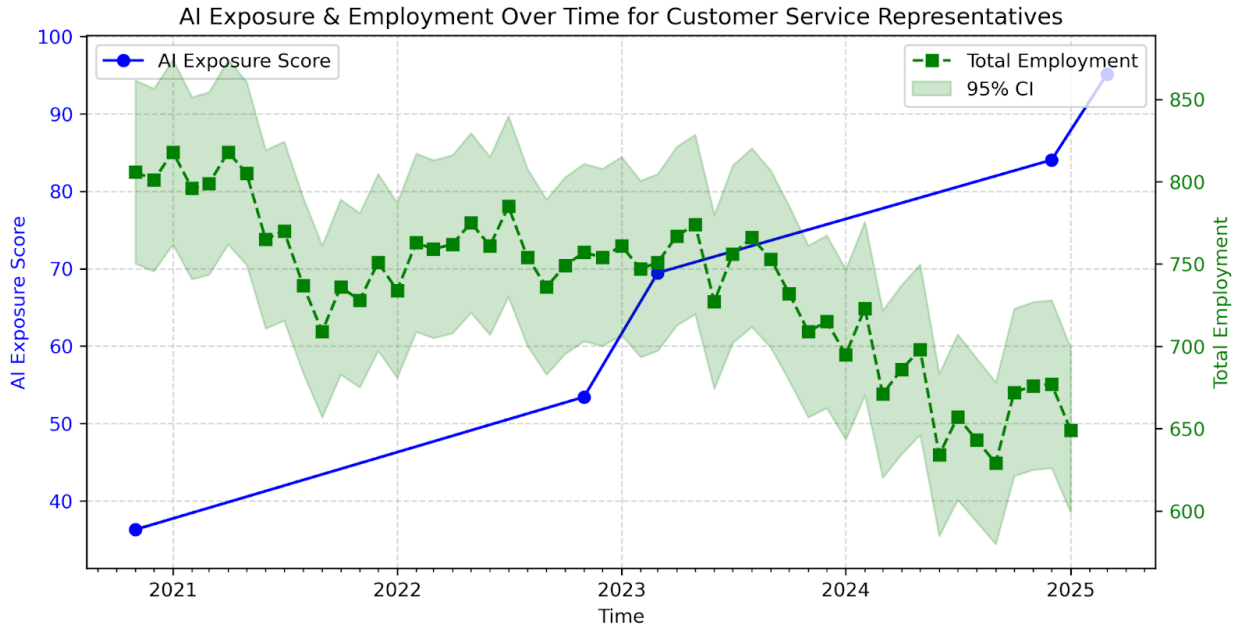
Figure 9. AI Exposure and Employment for Select Occupations

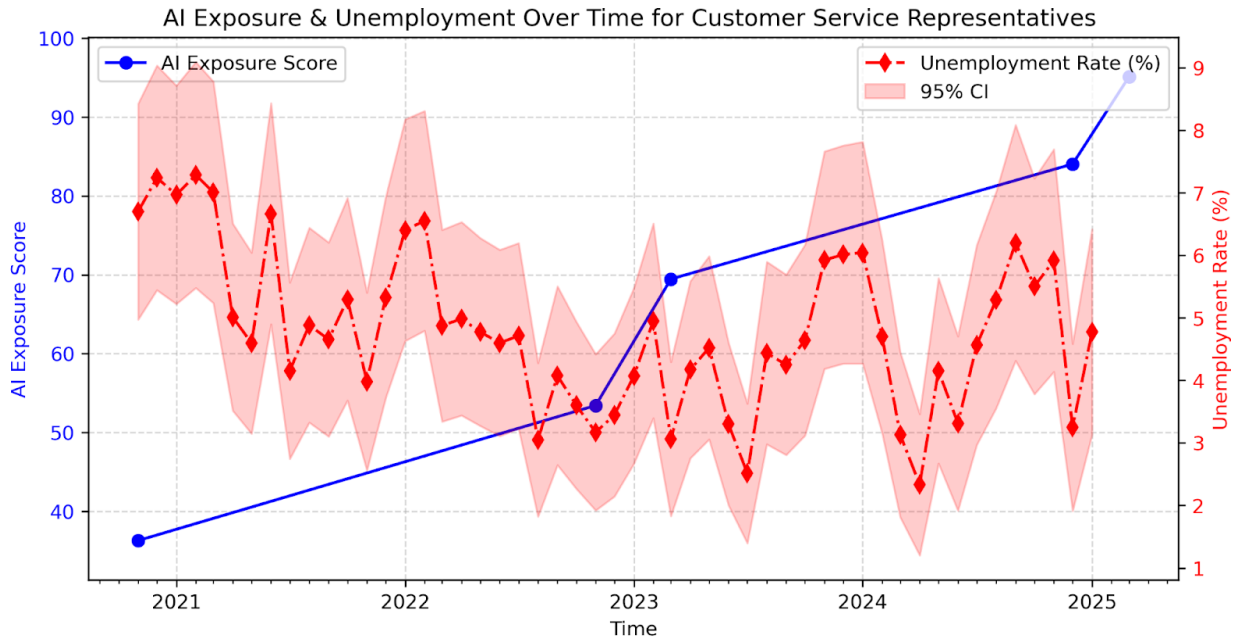
A. Software developers



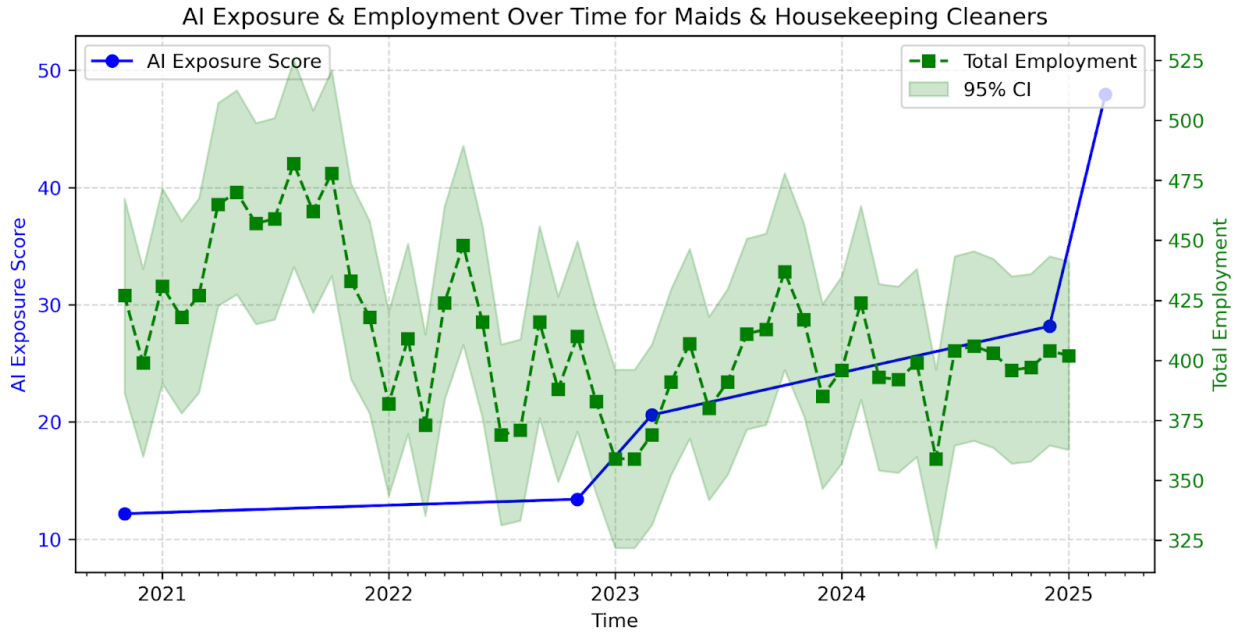


B. Customer Service Representatives





C. Maids and Housekeepers



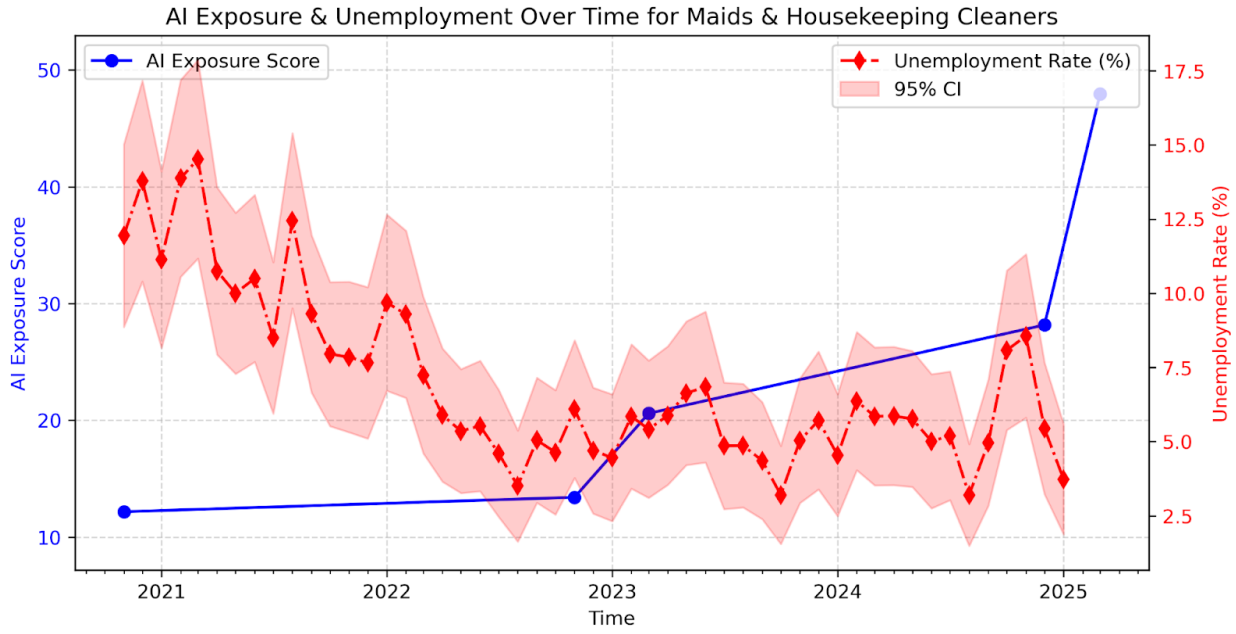


Figure 10. Dynamic treatment effects for all occupations: top quartile of Gen AI exposure vs. rest

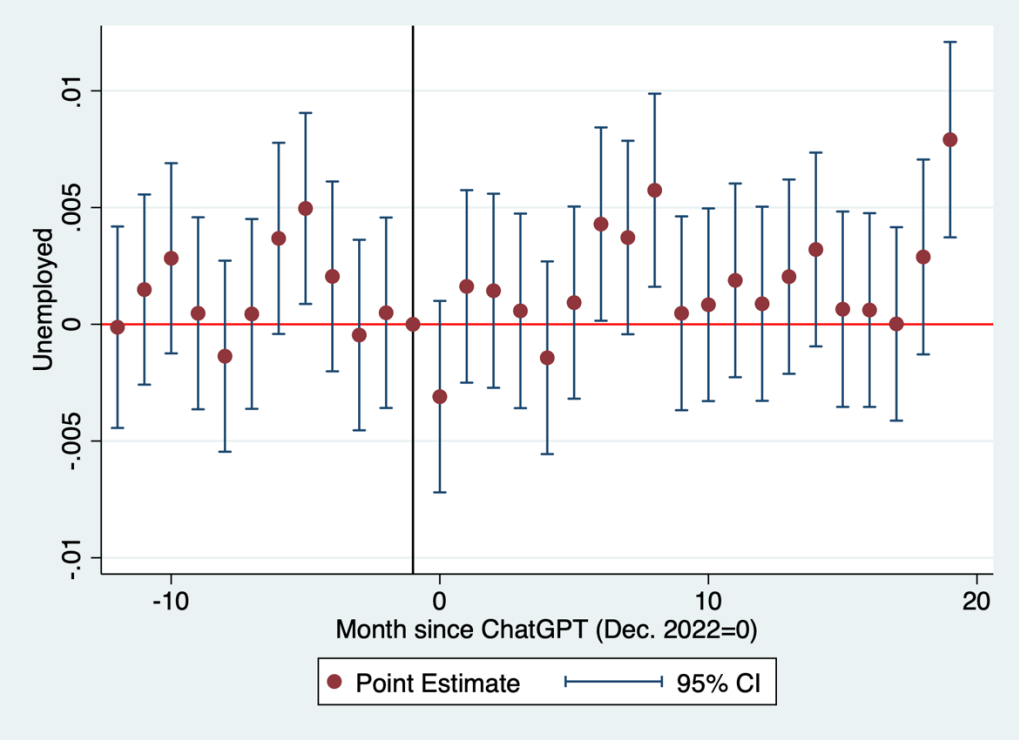
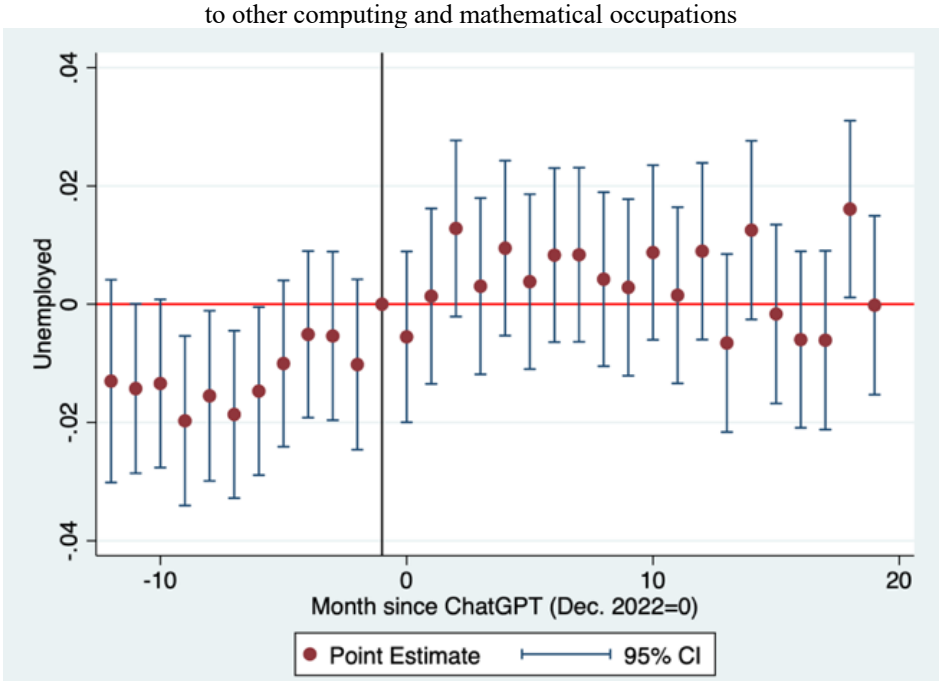


Figure 11. Dynamic treatment effects within Computer and Mathematical Occupations: Software developers relative to other computing and mathematical occupations



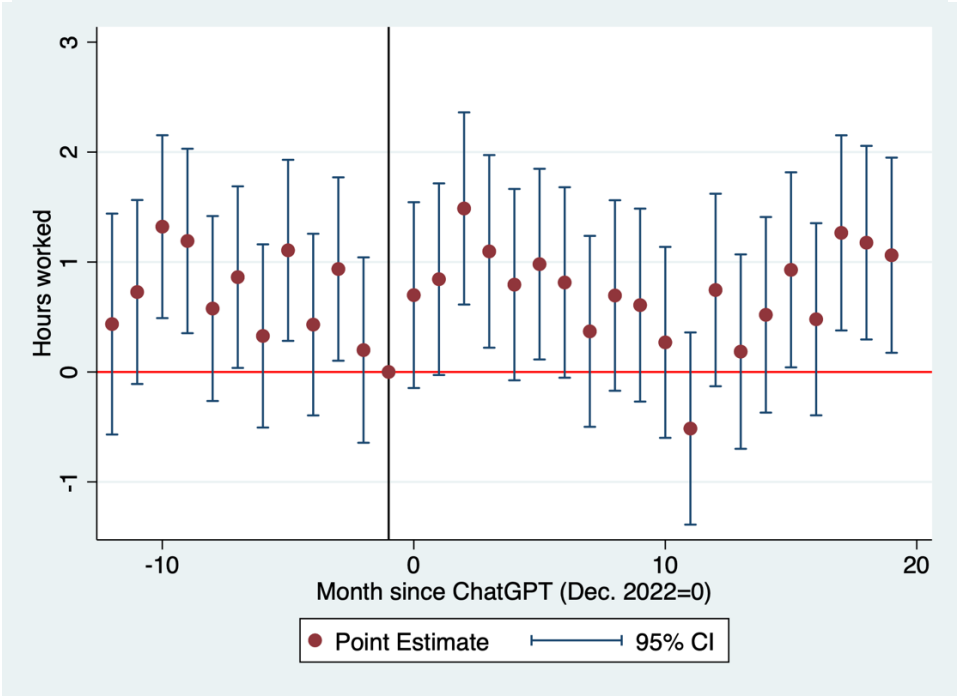


Figure 12. Dynamic treatment effects: Business and Financial Operations Occupations *relative to* Management Occupations

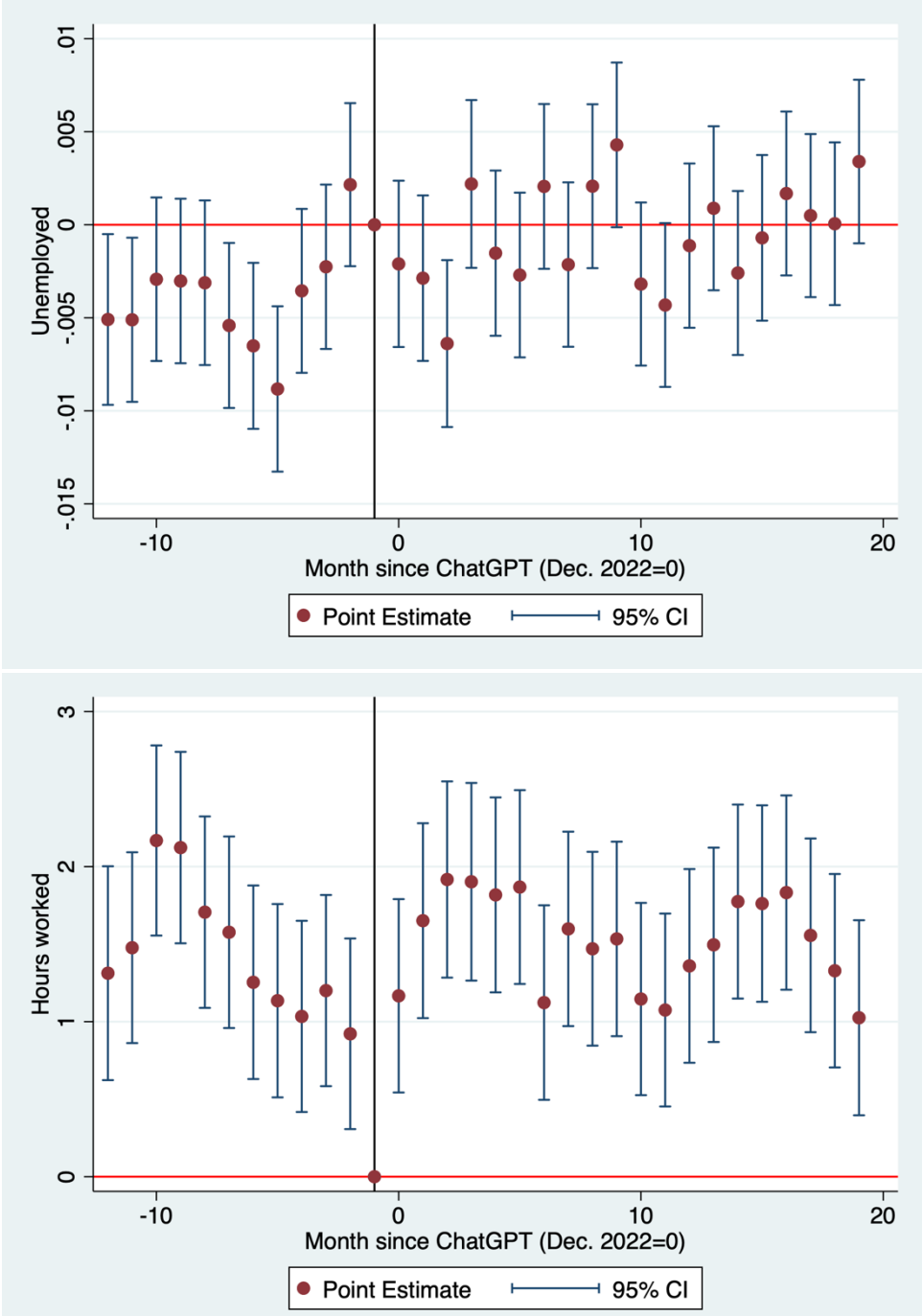


Table 1. Means and confidence levels by model and stage

Table 1: AI Exposure Scores by Model and Stage

Model	Stage	Mean Exposure	SD	Min	Max	High Conf.	Med. Conf.	Low Conf.
Claude	5	84.84	8.46	54.94	97.06	0.30	0.47	0.23
OpenAI	5	81.59	11.04	42.70	96.62	0.71	0.19	0.10
Claude	4	71.54	10.62	40.21	91.04	0.71	0.29	0.00
OpenAI	4	64.54	13.25	27.85	90.38	0.76	0.25	0.00
Claude	3	56.74	10.22	28.86	82.32	0.23	0.77	0.00
OpenAI	3	48.95	12.10	18.79	81.35	0.11	0.89	0.00
Llama	3	32.57	13.52	3.77	68.81			
Claude	2	43.22	10.71	16.66	75.37	0.30	0.70	0.00
OpenAI	2	34.74	11.49	10.10	73.01	0.18	0.82	0.01
Llama	2	18.40	8.27	1.31	42.54			
ChatGPT	2	19.38	13.35	2.63	59.88			
Claude	1	30.76	9.74	9.88	67.12	0.59	0.41	0.00
OpenAI	1	25.13	10.36	7.32	64.55	0.24	0.75	0.01
Llama	1	15.70	6.25	1.32	37.46			

Table 2. Top 40 Occupations by AI Exposure and Stage (Open AI Scores)

Stage 1

Rank	Occupation	AI Exposure Score
1	Accounting Clerks	64.55
2	Tellers	60.51
3	Billing Clerks	58.10
4	Typists	56.75
5	Order Clerks	56.50
6	Credit Clerks	56.16
7	Data Entry Keyers	55.46
8	Statistical Assistants	55.43
9	Payroll Clerks	54.41
10	Recordkeepers	52.49
11	Packaging Machine Ops.	52.47
12	Agricultural Sorters	51.98
13	Credit Analysts	50.79
14	Shipping Clerks	49.83
15	Inspectors	49.62
16	File Clerks	48.82
17	Insurance Clerks	48.24
18	Loan Interviewers	47.47
19	Photograph Processors	46.54
20	Packers	46.24
21	Cashiers	45.94
22	New Accounts Clerks	45.44
23	Cargo Agents	44.98
24	Machine Feeders	44.80
25	Medical Record Specialists	44.71
26	Polishing & Mixing Workers	44.36
27	Record Clerks, Other	44.34
28	Procurement Clerks	44.33
29	Travel Agents	44.13
30	Production Workers	43.48
31	Telephone Operators	43.46
32	Mail Clerks	43.44
33	Stockers	43.42
34	Receptionists	43.27
35	Postal Clerks	42.89
36	Mail Carriers	42.76
37	Production Clerks	42.65
38	Prepress Technicians	42.50
39	Office Machine Ops.	42.35
40	Adhesive Machine Ops.	42.07

Stage 3

Rank	Occupation	AI Exposure Score
1	Accounting Clerks	81.35
2	Tellers	80.22
3	Typists	77.74
4	Credit Clerks	77.45
5	Order Clerks	77.14
6	Billing Clerks	76.68
7	Payroll Clerks	76.26
8	Statistical Assistants	76.22
9	Insurance Clerks	75.79
10	Telemarketers	74.15
11	Travel Agents	72.65
12	Medical Secretaries	71.88
13	Data Entry Keyers	71.67
14	Credit Analysts	71.67
15	Switchboard Operators	71.58
16	Loan Interviewers	71.24
17	Medical Records Specialists	70.88
18	Shipping Clerks	70.62
19	Receptionists	70.56
20	Agricultural Sorters	70.56
21	Telephone Operators	70.49
22	Record Clerks, Other	70.29
23	Rental Clerks	70.23
24	File Clerks	70.18
25	Procurement Clerks	69.98
26	New Accounts Clerks	69.61
27	Customer Service Reps	69.46
28	Cashiers	69.31
29	General Secretaries	69.17
30	Dispatchers, Non-Emergency	68.69
31	Inspectors	68.41
32	Prepress Technicians	68.26
33	Recordkeepers	67.92
34	Credit Counselors	67.07
35	Cargo Agents	66.78
36	Photo Processors	66.69
37	Production Clerks	66.41
38	Office Clerks, Gen.	66.39
39	Medical Transcriptionists	66.08
40	Travel Clerks	65.98

Stage 5

Rank	Occupation	AI Exposure Score
1	Accounting Clerks	96.62
2	Payroll Clerks	96.54
3	Tellers	96.42
4	Credit Clerks	96.36
5	Insurance Clerks	96.07
6	Order Clerks	96.07
7	Billing Clerks	95.84
8	Telemarketers	95.76
9	Statistical Assistants	95.74
10	Typists	95.67
11	Medical Secretaries	95.42
12	Customer Service Reps	95.10
13	Record Clerks, Other	95.10
14	Loan Interviewers	95.07
15	Medical Transcriptionists	95.02
16	Travel Agents	95.00
17	Medical Records Specialists	94.71
18	Dispatchers, Non-Emergency	94.62
19	Telephone Operators	94.59
20	Rental Clerks	94.49
21	HR Assistants (Non-Payroll)	94.45
22	Data Entry Keyers	94.26
23	Credit Analysts	94.24
24	Proofreaders	94.20
25	Tax Preparers	94.17
26	Switchboard Operators	94.08
27	Procurement Clerks	93.91
28	Public Safety Telecommunicators	93.89
29	Court Clerks	93.63
30	Insurance Underwriters	93.61
31	Credit Counselors	93.60
32	Court Reporters	93.48
33	Shipping Clerks	93.19
34	Securities	93.04
35	Legal Secretaries	93.03
36	Government Eligibility Clerks	92.98
37	Prepress Technicians	92.94
38	Executive Secretaries	92.94
39	General Secretaries	92.92
40	Production Clerks	92.85

Table 3. Bottom 40 Occupations by AI Exposure and Stage (Open AI Scores)

Stage 1

Rank	Occupation	AI Exp. Score
1	Ambulance drivers	7.32
2	Actors	7.60
3	Drywall installers	7.85
4	Roofers	8.71
5	Musicians	8.76
6	Dentists	9.08
7	Dancers	9.12
8	Paramedics	9.23
9	Fence erectors	9.28
10	Groundskeepers	9.44
11	Iron/steel workers	9.68
12	Tree trimmers	9.69
13	Athletes	10.00
14	Stucco masons	10.28
15	Clergy	10.49
16	Firefighters	10.62
17	EMTs	10.83
18	Funeral attendants	10.87
19	Optometrists	11.00
20	Janitors	11.01
21	Lawyers	11.04
22	Power-line workers	11.26
23	Nurse anesthetists	11.28
24	Construction helpers	11.44
25	Telecom installers	11.55
26	Psychologists	11.56
27	Psychiatric aides	11.65
28	Interior designers	12.02
29	Auto body repairers	12.13
30	Housekeepers	12.17
31	Massage therapists	12.19
32	Cement finishers	12.21
33	Millwrights	12.26
34	Tile installers	12.39
35	Psychiatric techs	12.39
36	Manicurists	12.54
37	Surgeons	12.59
38	Rec therapists	12.67
39	Glaziers	12.70
40	Painters	12.72

Stage 3

Rank	Occupation	AI Exp. Score
1	Drywall installers	18.79
2	Ambulance drivers	19.57
3	Fence erectors	19.63
4	Housekeepers	20.59
5	Roofers	20.69
6	Tree trimmers	21.29
7	Construction helpers	21.52
8	Iron/steel workers	22.07
9	Telecom installers	22.53
10	Groundskeepers	22.58
11	Janitors	22.68
12	Stucco masons	23.47
13	Cement finishers	25.23
14	Power-line workers	25.28
15	Millwrights	25.29
16	Glaziers	25.61
17	Dishwashers	25.78
18	Tile installers	25.88
19	Auto body repairers	25.91
20	Cafeteria attendants	25.99
21	Painters	26.48
22	Funeral attendants	26.50
23	Psychiatric aides	26.80
24	Firefighters	26.90
25	Manicurists	26.94
26	Highway workers	27.67
27	Oil/gas drillers	27.83
28	Brickmasons	28.43
29	Rail-track workers	28.45
30	Dentists	28.49
31	Insulation workers	29.11
32	Construction laborers	29.33
33	Carpenters	29.79
34	Riggers	29.89
35	Vehicle mechanics	29.99
36	Paramedics	30.00
37	Food prep workers	30.06
38	Auto glass repairers	30.23
39	Vehicle cleaners	30.24
40	Plumbers	30.45

Stage 5

Rank	Occupation	AI Exp. Score
1	Roofers	28.95
2	Drywall installers	27.85
3	Telecom installers	30.36
4	Fence erectors	28.28
5	Tree trimmers	29.44
6	Housekeepers	28.17
7	Stucco masons	33.13
8	Ambulance drivers	28.72
9	Manicurists	35.00
10	Construction helpers	29.97
11	Locksmiths	36.48
12	Painters	35.56
13	Tile installers	35.25
14	Janitors	31.01
15	Hairdressers	42.21
16	Groundskeepers	31.64
17	Glaziers	36.07
18	Iron/steel workers	32.37
19	Funeral attendants	38.09
20	Power-line workers	36.40
21	Cement finishers	36.03
22	Brickmasons	38.53
23	Insulation workers	38.84
24	Carpenters	39.96
25	Cafeteria attendants	31.80
26	Barbers	42.27
27	Firefighters	39.91
28	Psychiatric aides	37.65
29	Extraction workers	40.83
30	Vehicle mechanics	38.45
31	Dishwashers	33.24
32	Auto body repairers	37.33
33	Millwrights	36.75
34	Dentists	43.81
35	Skincare specialists	46.56
36	Construction workers	43.03
37	Plumbers	42.97
38	Auto glass repairers	40.56
39	Nursing assistants	44.14
40	Truck mechanics	44.22

Table 4. Stage 2 AI capabilities and changes in worker outcomes (3-month averages: Sep-Nov 2022 to Sep-Nov 2024)

	(1) Employed	(2) Unemp. Rate	(3) Hours Worked - Main Job	(4) Hours Worked - All Jobs	(5) Employed	(6) Unemp. Rate	(7) Hours Worked - Main Job	(8) Hours Worked - All Jobs
Panel A. No Controls								
ChatGPT Exposure	-0.104***	3.96e-05	0.00228	0.0102***				
	(0.0103)	(3.22e-05)	(0.00179)	(0.00205)				
Claude Exposure					-0.0799***	0.000108***	-0.00368**	0.00680***
					(0.00943)	(3.39e-05)	(0.00177)	(0.00197)
Observations	19,387	19,387	19,350	19,350	19,387	19,387	19,350	19,350
R-squared	0.006	0.000	0.000	0.001	0.004	0.001	0.000	0.001
Panel B. With Gender, Education, Age, Race, and Regional Controls								
ChatGPT Exposure	-0.887***	0.000383***	-0.0108***	-0.00781***				
	(0.0711)	(2.81e-05)	(0.00191)	(0.00211)				
Claude Exposure					-0.988***	0.000455***	-0.0134***	-0.0115***
					(0.0827)	(3.22e-05)	(0.00213)	(0.00231)
Observations	2,356	2,356	2,356	2,356	2,356	2,356	2,356	2,356
R-squared	0.250	0.305	0.199	0.199	0.248	0.311	0.202	0.204

Table 5. Stage 2 AI capabilities and changes in worker outcomes (6-month averages: Jun-Nov 2022 to Jun-Nov 2024)

	(1) Employed	(2) Unemp. Rate	(3) Hours Worked - Main Job	(4) Hours Worked - All Jobs	(5) Employed	(6) Unemp. Rate	(7) Hours Worked - Main Job	(8) Hours Worked - All Jobs
Panel A. No Controls								
ChatGPT Exposure	-0.114***	8.08e-05***	0.00326**	0.0105***				
	(0.00923)	(2.77e-05)	(0.00146)	(0.00176)				
Claude Exposure					-0.0925***	0.000130***	-0.000836	0.00700***
					(0.00841)	(2.94e-05)	(0.00151)	(0.00174)
Observations	19,406	19,406	19,406	19,406	19,406	19,406	19,406	19,406
R-squared	0.009	0.000	0.000	0.002	0.006	0.001	0.000	0.001
Panel B. With Gender, Education, Age, Race, and Regional Controls								
ChatGPT Exposure	-0.443***	0.000268***	-0.00129	0.00256**				
	(0.0535)	(1.86e-05)	(0.00112)	(0.00122)				
Claude Exposure					-0.614***	0.000302***	0.000205	0.00274*
					(0.0624)	(2.18e-05)	(0.00147)	(0.00146)
Observations	3,382	3,382	3,382	3,382	3,382	3,382	3,382	3,382
R-squared	0.176	0.248	0.224	0.221	0.185	0.248	0.224	0.221

Table 6. Stage 3 AI capabilities and changes in worker outcomes (3-month averages: Sep-Nov 2022 to Sep-Nov 2024)

	(1) Employed	(2) Unemp. Rate	(3) Hours Worked - Main Job	(4) Hours Worked - All Jobs	(5) Employed	(6) Unemp. Rate	(7) Hours Worked - Main Job	(8) Hours Worked - All Jobs
Panel A. No Controls								
ChatGPT Exposure	-0.0941***	6.94e-05**	-0.00175	0.00743***				
	(0.00994)	(3.03e-05)	(0.00173)	(0.00196)				
Claude Exposure					-0.0720***	0.000150***	-0.00508***	0.00829***
					(0.00996)	(3.62e-05)	(0.00192)	(0.00215)
Observations	19,387	19,387	19,350	19,350	19,387	19,387	19,350	19,350
R-squared	0.005	0.000	0.000	0.001	0.003	0.001	0.000	0.001
Panel B. With Gender, Education, Age, Race, and Regional Controls								
ChatGPT Exposure	-0.920***	0.000421***	-0.0112***	-0.00724***				
	(0.0604)	(2.53e-05)	(0.00186)	(0.00213)				
Claude Exposure					-1.169***	0.000521***	-0.0187***	-0.0154***
					(0.0818)	(3.44e-05)	(0.00225)	(0.00249)
Observations	2,356	2,356	2,356	2,356	2,356	2,356	2,356	2,356
R-squared	0.258	0.324	0.201	0.198	0.257	0.319	0.214	0.211

Table 7. Stage 3 AI capabilities and changes in worker outcomes (6-month averages: Jun-Nov 2022 to Jun-Nov 2024)

	(1) Employed	(2) Unemp. Rate	(3) Hours Worked - Main Job	(4) Hours Worked - All Jobs	(5) Employed	(6) Unemp. Rate	(7) Hours Worked - Main Job	(8) Hours Worked - All Jobs
Panel A. No Controls								
ChatGPT Exposure	-0.106***	8.97e-05***	0.00292**	0.0109***				
	(0.00928)	(2.61e-05)	(0.00139)	(0.00170)				
Claude Exposure					-0.0880***	0.000169***	0.00152	0.0116***
					(0.00917)	(3.06e-05)	(0.00155)	(0.00186)
Observations	19,406	19,406	19,406	19,406	19,406	19,406	19,406	19,406
R-squared	0.008	0.001	0.000	0.002	0.005	0.002	0.000	0.002
Panel B. With Gender, Education, Age, Race, and Regional Controls								
ChatGPT Exposure	-0.439***	0.000334***	-0.00528***	0.000944				
	(0.0551)	(1.83e-05)	(0.00109)	(0.00122)				
Claude Exposure					-0.541***	0.000396***	-0.00689***	-2.98e-05
					(0.0678)	(2.38e-05)	(0.00133)	(0.00150)
Observations	3,382	3,382	3,382	3,382	3,382	3,382	3,382	3,382
R-squared	0.177	0.277	0.229	0.220	0.176	0.270	0.229	0.220

Table 8. Change in AI capabilities (Stage 1 to 3) and changes in worker outcomes (3-month averages: Sep-Nov 2022 to Sep-Nov 2024)

	(1) Employed	(2) Unemp. Rate	(3) Hours Worked - Main Job	(4) Hours Worked - All Jobs	(5) Employed	(6) Unemp. Rate	(7) Hours Worked - Main Job	(8) Hours Worked - All Jobs
Panel A. No Controls								
Change in ChatGPT Exposure (Stage 3 - Stage 1)	-0.0757***	0.0000943	-0.0329***	-0.0138***				
	(0.0167)	(6.41e-05)	(0.00336)	(0.00355)				
Change in Claude Exposure (Stage 3 - Stage 1)					-0.0459***	0.0000919	-0.0286***	-0.0102***
					(0.0166)	(6.76e-05)	(0.00340)	(0.00357)
Observations	19,387	19,387	19,350	19,350	19,387	19,387	19,350	19,350
R-squared	0.001	0.000	0.005	0.001	0.000	0.000	0.004	0.000
Panel B. With Gender, Education, Age, Race, and Regional Controls								
Change in ChatGPT Exposure (Stage 3 - Stage 1)	-0.0664***	0.000117**	-0.0242***	-0.0126***				
	(0.0174)	(5.63e-05)	(0.00326)	(0.00343)				
Change in Claude Exposure (Stage 3 - Stage 1)					-0.0365**	0.000117*	-0.0221***	-0.0100***
					(0.0173)	(6.09e-05)	(0.00325)	(0.00344)
Observations	19,387	19,387	19,350	19,350	19,387	19,387	19,350	19,350
R-squared	0.004	0.024	0.075	0.061	0.004	0.024	0.075	0.060

Table 9. Difference-in-difference estimates

VARIABLES	(1) Unemployed	(2) Hours worked - main job	(3) Unemployed	(4) Hours worked - main job	(5) Unemployed	(6) Hours worked - main job	(7) Unemployed	(8) Hours worked - main job
Post x Top quartile of AI Exposure (Stage 1)	0.00172**	0.0186			0.00197***		-0.0101	
	(0.000730)	(0.0432)			(0.000728)		(0.0425)	
Post x Top quartile of AI Exposure (Stage 3)			0.00182**	-0.0491		0.00195***		-0.0737*
			(0.000712)	(0.0429)		(0.000710)		(0.0423)
Occupation FEs	Y	Y	Y	Y	Y	Y	Y	Y
Month-Year FEs	Y	Y	Y	Y	Y	Y	Y	Y
Additional controls					Y	Y	Y	Y
Observations	1,677,325	1,551,299	1,677,325	1,551,299	1,677,325	1,677,325	1,551,299	1,551,299
R-squared	0.014	0.112	0.014	0.112	0.019	0.019	0.133	0.133

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10. Heterogeneous effect – DD estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Unemployed				
VARIABLES	HS or less	BA or more	More than BA	30 or less	31 to 50	more than 50	male	female
Panel A. Stage 1 exposure								
Post x Top quartile of AI Exposure (Stage 1)	0.00419***	-0.00225**	0.00196	0.00474***	0.000538	0.000762	0.00108	0.00271***
	(0.00136)	(0.00111)	(0.00225)	(0.00169)	(0.00104)	(0.00114)	(0.00108)	(0.000997)
Observations	558,299	679,780	251,057	411,743	788,426	477,156	877,548	799,775
R-squared	0.014	0.011	0.014	0.017	0.014	0.011	0.016	0.014
Panel B. Stage 3 exposure								
Post x Top quartile of AI Exposure (Stage 3)	0.00261*	0.00103	0.00796***	0.00453***	0.000430	0.00126	0.00213**	0.00206**
	(0.00138)	(0.00104)	(0.00207)	(0.00167)	(0.001000)	(0.00114)	(0.00106)	(0.000966)
Observations	558,299	679,780	251,057	411,743	788,426	477,156	877,548	799,775
R-squared	0.014	0.011	0.014	0.017	0.014	0.011	0.016	0.014

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix: System Prompt – multi-stage cumulative capabilities (Open AI Chat GPT 4o and Anthropic Claude 3.5 Sonnet)

You are a helpful coding assistant.
I have a CSV file containing a list of tasks from various occupations in the economy. For each task in the CSV, I want to assess how well an AI system can perform that task across five distinct "stages" of AI model capabilities. Each stage builds cumulatively upon the prior stage, meaning higher-level models retain and expand upon the abilities of lower-level models. These stages are defined in detail below:
1. **Stage 1 – Traditional Machine Learning (pre-LLM)**
  - Uses classical machine learning techniques such as supervised learning, regression, clustering, and rule-based models.
  - Can process structured and semi-structured data for well-defined problems (e.g., fraud detection, sales forecasting).
  - Basic computer vision capabilities (e.g., image classification, object detection) using convolutional neural networks (CNNs).
  - Natural language understanding is weak—primarily relies on keyword matching, statistical models, or rule-based processing.
  - Cannot generate original text, reason through complex problems, or dynamically adapt to new situations.
  - Performs well on highly repetitive, structured tasks but lacks adaptability to novel inputs.
  - **Example Applications:**
    - Finance: Credit scoring, fraud detection.
    - Retail: Demand forecasting, customer segmentation.
    - Healthcare: Identifying tumors in medical images (basic classification).
    - Manufacturing: Defect detection in product images.
    - Logistics: Route optimization for delivery networks.

2. **Stage 2 – Early Large Language Models (text-based generative AI)**
  - Expands on Stage 1 by introducing text-based generative capabilities.
  - Can generate, summarize, and paraphrase text with moderate coherence.
  - Handles simple instruction-following and basic question-answering but struggles with deep reasoning.
  - Still single-modality (text-only); does not process images, audio, or video.
  - Factual accuracy is inconsistent; prone to hallucinations and misinterpretation of nuanced queries.
  - Struggles with multi-step reasoning, decision-making, and domain-specific expertise.
  - **Example Applications:**
    - Customer Support: Basic chatbots for FAQ responses.
    - Education: Automated essay grading based on keyword matching.
    - Legal: Extracting key terms from contracts.

- Writing & Editing: AI-assisted grammar correction.
- HR & Recruiting: Screening resumes based on keyword matching.

3. **Stage 3 – Multi-Modal Generative AI**
- Builds upon Stage 2 by introducing multi-modal capabilities (i.e., processing both text and images/audio).
- Improved context awareness, allowing it to integrate and analyze different input types.
- More advanced reasoning than early LLMs, but still prone to errors.
- Capable of image and audio processing but lacks deep analytical capability.
- Useful for tasks requiring both text and images/audio.
- **Example Applications:**
  - Medical: Generating radiology reports from X-rays or MRIs.
  - Design: Automatically generating product mockups from textual descriptions.
  - Media & Journalism: Generating audio transcripts and summarizing news.
  - Retail & E-commerce: Product image tagging and personalized recommendations.
  - Finance: Analyzing charts and tables for market trends.

4. **Stage 4 – Advanced Reasoning Models**
- Expands on Stage 3 by significantly improving logical reasoning and problem-solving.
- Capable of multi-step reasoning, structured analysis, and more reliable fact-based responses.
- Better at specialized and domain-specific queries (e.g., finance, law, medicine).
- Reduces hallucinations and can cross-check factual consistency.
- Still predominantly text-based but far superior in analytical reasoning.
- **Example Applications:**
  - Medicine: AI-assisted diagnosis based on symptoms and lab results.
  - Programming: Writing and debugging complex code.
  - Finance: Conducting risk analysis and market forecasting.
  - Legal: Analyzing and summarizing legal cases.
  - Engineering: Simulating structural loads and materials behavior.

5. **Stage 5 – Autonomous AI Agent Models**
- Builds on Stage 4 by introducing autonomous action-taking.
- Can interact with software environments, call APIs, execute code, and manage files.
- Plans and executes multi-step workflows with minimal human input.
- Capable of retrieving external data, leveraging tools, or collaborating with plugins to extend functionality.
- Maintains context across iterative steps, can revise actions based on feedback or observed outcomes.
- **Example Applications:**

- Scientific Research: Conducting autonomous experiments, analyzing results, and refining hypotheses.
- Business & Finance: Fully automating market research, competitor analysis, and trading strategies.
- Software Development: Writing, testing, and deploying software autonomously.
- Healthcare: Running AI-assisted drug discovery processes.
- Cybersecurity: Monitoring systems and autonomously patching security vulnerabilities.

For each row in my CSV (each row containing a specific task description), please do the following **for each of the five stages**:

1. **Estimate the percentage (0–100%)** of how much of that task an AI system at this stage could reasonably complete or assist with.
2. **Explain your rationale** (a concise reason).
3. **State your confidence level** in your assessment (High, Medium, or Low).

Return your results in a structured JSON format. The JSON should have a top-level key "stages", with sub-keys "Stage 1" through "Stage 5". Each stage should have keys "percentage", "rationale", and "confidence".

Appendix: Sample Prompt – point in time (Meta Llama 3.0 70b)

Consider State-Of-The-Art generative AI (Gen AI) technologies currently available, including language models, image generators, and video generators. Acknowledge that Gen AI has significantly evolved since earlier large language models (LLMs). These technologies are now capable of completing a wide range of tasks involving text, image, video, or structured data inputs and outputs. Over time, Gen AI has improved in its ability to handle tasks more efficiently and accurately across multiple formats. Current State-Of-The-Art Gen AI models have enhanced their ability to reduce the time needed to complete tasks through improved reasoning, contextual understanding, and multimodal capabilities. These models can also process more complex and diverse tasks compared to earlier versions. Unlike earlier models, State-Of-The-Art Gen AI can access real-time information from external sources through APIs, reducing the need for pre-provided context and enabling faster and more accurate task completion. State-Of-The-Art Gen AI technologies can be fine-tuned for specific industries or domains, allowing for greater task precision and efficiency in specialized areas. State-Of-The-Art Gen AI can process and generate not just text but also images, video, and structured data (e.g., tables, charts, PDFs), making these technologies more versatile for a wide range of tasks across multiple formats. \n

You are a helpful research assistant with a deep understanding of the how AI technologies have evolved in the above manner and who wants to assess the given task according to the instructions below. \n

Assume a worker with average expertise in that role. The worker has access to State-Of-The-Art generative AI technologies described above as well as any other existing software or computer hardware tools mentioned in the task. The worker has access to any commonly available technical tools accessible via a laptop (e.g. a microphone, speakers, etc.). The worker does not have access to any other physical tools or materials. \n

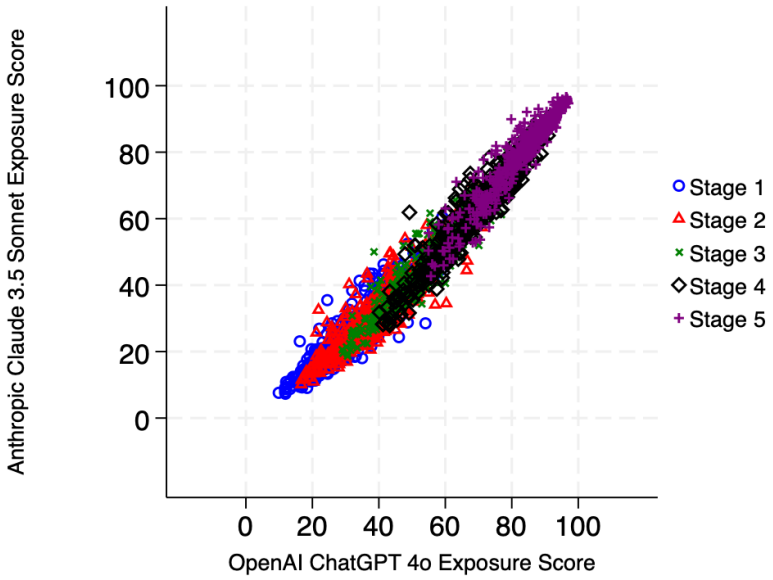
Read the following occupation title and one of the tasks that belong to that occupation. Then do three things. \n

1: Reason step by step to answer what percent of each task can be performed by current State-Of-The-Art generative AI tools. If you aren't sure how to judge what percent of the task can be performed by generative AI, consider the subtasks associated with the task and assess what percentage of all subtasks can be performed by State-Of-The-Art generative AI tools. Your answer must range from 0% to 100%. You can round up to the nearest 5%. Give an explanation for your answer. Your answer here should start with 1: followed by the description. \n

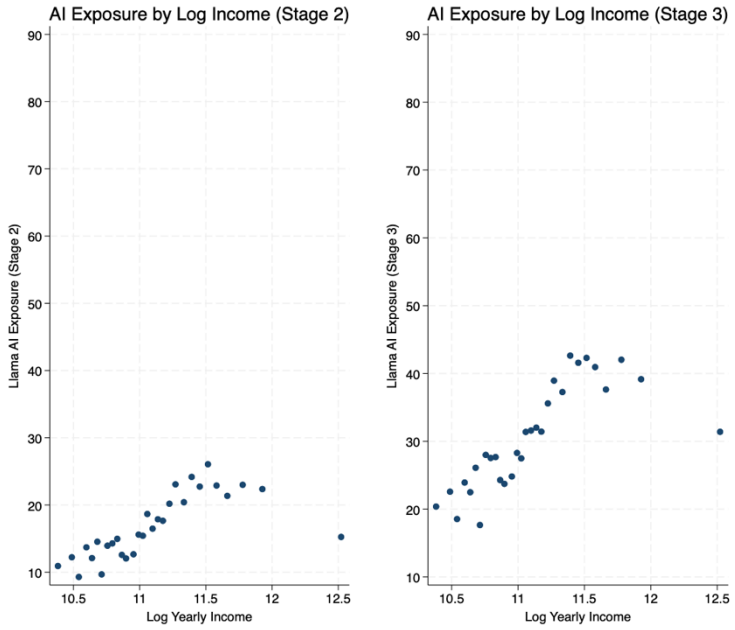
2: Report only the percent that you determined for the task, which should match your answer in step 1. Do not reply N/A. Your answer here should start with 2: followed by the percent. \n

3: Given the amount of speculation required in step 1, describe your certainty about the estimate–either high, moderate, or low. Your answer here should start with 3: followed by your uncertainty. \n

Appendix Figure 1. Correlation plot between ChatGPT 4o and Claude 3.5 Sonnett occupational AI exposure scores



Appendix Figure 2. Occupational AI exposure by income
-Based on Llama 3.0 assessment of Stage 2 and 3 capabilities.



Appendix Table 1. Model development stages

Stage	Start Date	Notes
1	Before Nov. 2022	Pre LLM AI capabilities focused on ML, NLP, CV, etc.
2	November 2022	GPT-3.5 was released in November 2022. Earlier models of GPT were released before this (GPT-2, GPT-3), however, GPT-3.5 was the first widely usable GPT.
3	March 2023	GPT-4 was released in March 2023 and was the first GPT model that allowed multimodal inputs.
4	December 2024	GPT-o1 was fully released to the public in December 2024. Before this, OpenAI released GPT-4o, which improved on speed and multimodal inputs but didn't include advanced reasoning. GPT-o1 had a limited release in September 2024, but this release was only available for a limited pool of users for early testing.
5	March 2025 – In development	Agentic AI that can perform actions on the computer are being introduced and continue to be developed