Future of Work with AI Agents:

Auditing Automation and Augmentation Potential across the U.S. Workforce

Yijia Shao*, Humishka Zope*, Yucheng Jiang, Jiaxin Pei, David Nguyen,
Erik Brynjolfsson, Diyi Yang
Stanford University
{shaoyj, diyiy}@cs.stanford.edu

Abstract

The rapid rise of compound AI systems (*a.k.a.*, AI agents) is reshaping the labor market, raising concerns about job displacement, diminished human agency, and overreliance on automation. Yet, we lack a systematic understanding of the evolving landscape. In this paper, we address this gap by introducing a novel auditing framework to assess which occupational tasks workers want AI agents to automate or augment, and how those desires align with the current technological capabilities. Our framework features an audio-enhanced mini-interview to capture nuanced worker desires and introduces the Human Agency Scale (HAS) as a shared language to quantify the preferred level of human involvement. Using this framework, we construct the WORKBank database, building on the U.S. Department of Labor's O*NET database, to capture preferences from 1,500 domain workers and capability assessments from AI experts across over 844 tasks spanning 104 occupations. Jointly considering the desire and technological capability divides tasks in WORKBank into four zones: Automation "Green Light" Zone, Automation "Red Light" Zone, R&D Opportunity Zone, Low Priority Zone. This highlights critical mismatches and opportunities for AI agent development. Moving beyond a simple automate-or-not dichotomy, our results reveal diverse HAS profiles across occupations, reflecting heterogeneous expectations for human involvement. Moreover, our study offers early signals of how AI agent integration may reshape the core human competencies, shifting from information-focused skills to interpersonal ones. These findings underscore the importance of aligning AI agent development with human desires and preparing workers for evolving workplace dynamics.

1 Introduction

Rapid advances in foundation models, such as large language models (LLMs), have catalyzed growing interest in AI agents: goal-directed systems equipped with tool access and multi-step execution capabilities. Unlike standalone models, these agents can perform complex workflows and are increasingly positioned to take on roles across a broad range of professional domains (Jiang et al., 2024, Shao et al., 2024a, Wang et al., 2024b, Yang et al., 2024, Yao et al., 2024). Their integration into occupational settings is already beginning to shape the labor market (Demirci et al., 2025, Hoffmann et al., 2024). For example, research indicates that around 80% of U.S. workers may see LLMs affect at least 10% of their tasks, with 19% potentially seeing over half impacted (Eloundou et al., 2023). Usage data from Anthropic indicates that in early 2025, at least some workers in 36% of occupations *already* were using AI for at least 25% of their tasks (Handa et al., 2025).

While AI adoption in the workplace has shown promise in boosting productivity, it also raises concerns about job displacement, reduced human agency, and overreliance on automation (Hazra et al.,

---

*Equal Contribution

2025). Despite this critical impact, we lack a systematic and grounded understanding of the evolving landscape. From a coverage perspective, prior research often focuses on a few domains like software engineering (Hoffmann et al., 2024) and customer support (Brynjolfsson et al., 2025). This narrow scope limits our comprehension of the real-world complexity of diverse human jobs and the varied nature of open-ended tasks. From a stakeholder perspective, existing studies often emphasize the interests of capital by focusing on a few tasks that tend to be more profitable such as coding without adequately considering worker values (Eisfeldt et al., 2023). Furthermore, current approaches often rely on analyzing existing usage data, such as how people use chatbots for work (Hazra et al., 2025, Zhao et al., 2024), which cannot provide a forward-looking assessment of AI potential across the broader workforce.

To address these gaps, we propose a principled, survey-based framework to investigate which occupational tasks workers want AI agents to automate or augment. We look at the entire workforce that could be impacted by digital AI agents by sourcing occupational tasks from the U.S. Department of Labor's O*NET database. Compared to occupation-level studies, task-level auditing allows us to better capture the nuanced, open-ended, and contextual nature of real-world work. Our auditing framework takes a worker-centric approach by soliciting first-hand insights from domain workers actively performing the tasks. To guide domain workers in providing well-calibrated responses, we empower them to share their experiences and articulate their reasoning through an audio-enhanced survey system. Crucially, our framework expands beyond the binary view of automation. We propose the Human Agency Scale (*i.e.*, H1-H5), which complements the SAE L0-L5 automation levels (Committee et al., 2014) by quantifying the degree of human involvement required for occupational task completion and quality. This new scale centers human agency—a crucial factor for responsible AI agent adoption (Fanni et al., 2023)—and provides a shared language to capture the spectrum between automation and augmentation.

To ground workers' perspectives in technical reality, we further gather complementary assessments from AI experts with experience in agent research and development (R&D). This dual approach reveals how workers and experts perceive AI agents' capabilities and risks at work. Based on data collected through January 2025 to May 2025, we construct the AI Agent **W**orker **O**utlook & **R**eadiness **K**nowledge Bank (WORKBank). This database currently consists of responses from 1,500 workers across 104 occupations and annotations from 52 AI experts, covering 844 occupational tasks. It is designed to be easily extensible to more tasks and to reflect evolving technological capabilities and worker preferences. To our knowledge, this is the first large-scale audit of AI agent capabilities and worker preferences.

Our work contributes four sets of findings:

1. **Domain workers want automation for low-value and repetitive tasks (Figure 4).** For 46.1% of tasks, workers express positive attitudes toward AI agent automation, even after reflecting on potential job loss concerns and work enjoyment. The primary motivation for automation is freeing up time for high-value work, though trends vary significantly by sector.

2. **We visualize the desire-capability landscape of AI agents at work, and find critical mismatches (Figure 5).** The worker desire and technological capability divide the landscape into four zones: Automation "Green Light" Zone (high desire and capability), Automation "Red Light" Zone (high capability but low desire), R&D Opportunity Zone (high desire but currently low capability), and Low Priority Zone (low desire and low capability). Notably, 41.0% of Y Combinator company-task mappings are concentrated in the Low Priority Zone and Automation "Red Light" Zone. Current investments mainly center around software development and business analysis, leaving many promising tasks within the "Green Light" Zone and Opportunity Zone under-addressed.

3. **The Human Agency Scale provides a shared language to audit AI use at work and reveals distinct patterns across occupations (Figure 6).** 45.2% of occupations have H3 (equal partnership) as the dominant worker-desired level, underscoring the potential for human-agent collaboration. However, workers generally prefer higher levels of human agency, potentially foreshadowing friction as AI capabilities advance.

4. **Key human skills are shifting from information processing to interpersonal competence (Figure 7).** By mapping tasks to core skills and comparing their associated wages and required human agency, we find that traditionally high-wage skills like analyzing information are becoming less emphasized, while interpersonal and organizational skills are gaining more importance. Additionally, there is a trend toward requiring broader skill sets from individuals. These patterns offer early signals of how AI agent integration may reshape core competencies.

2 Auditing Framework

To investigate how AI agents may integrate into professional work, we develop a task-level, survey-based auditing framework that captures both worker preferences and technological feasibility across the automation–augmentation spectrum. We begin by outlining a few key design principles of our framework before examining each one in detail.

2.1 Defining Audit Granularity and Scope

Our framework focuses on *complex*, *multi-step tasks* associated with specific occupations (*e.g.*, "Marketing Managers: Compile lists describing product or service offerings"), sourced from the O*NET database. These tasks, unlike isolated, low-level activities (*e.g.*, "track goods or materials" or "translate information"), reflect actual job responsibilities and the kinds of workflows AI agents are poised to impact. Moreover, compared with occupation-level analysis, conducting the audit at the task level enables a more nuanced understanding, as tasks within the same profession can vary significantly and are often highly contextualized.

We scope our audit to computer-compatible tasks, recognizing their susceptibility to *foundation model-powered AI agents*. Drawing from historical and recent literature on agent autonomy (Castelfranchi, 1994, Russell and Norvig, 1995, Wooldridge and Jennings, 1995), planning capabilities, and tool use (Mitchell et al., 2025), we define AI agents (excluding physical robots) as: *"A system or program capable of autonomously performing tasks on behalf of a user or another system by designing its workflow and utilizing available software tools, without the ability to perform physical actions."*

2.2 Emphasizing the Spectrum of Automation and Augmentation

Traditional technology impact studies often ask: *To what degree can this task be automated?* Besides this view of automation, we consider the view of augmentation—where technology complements and enhances human capabilities (Autor, 2015), as this new wave of technology holds significant potential to augment human workers through human-agent collaboration, enhancing both productivity and work quality. While augmentation has been discussed in prior work (Brynjolfsson, 2022, Handa et al., 2025), there is no established framework for quantifying automation vs. augmentation. To fill this gap and provide a shared language, we introduce the Human Agency Scale (HAS) (Figure 2), a five-level scale from H1 (no human involvement) to H5 (human involvement essential):

- H1: AI agent handles the task entirely on its own.
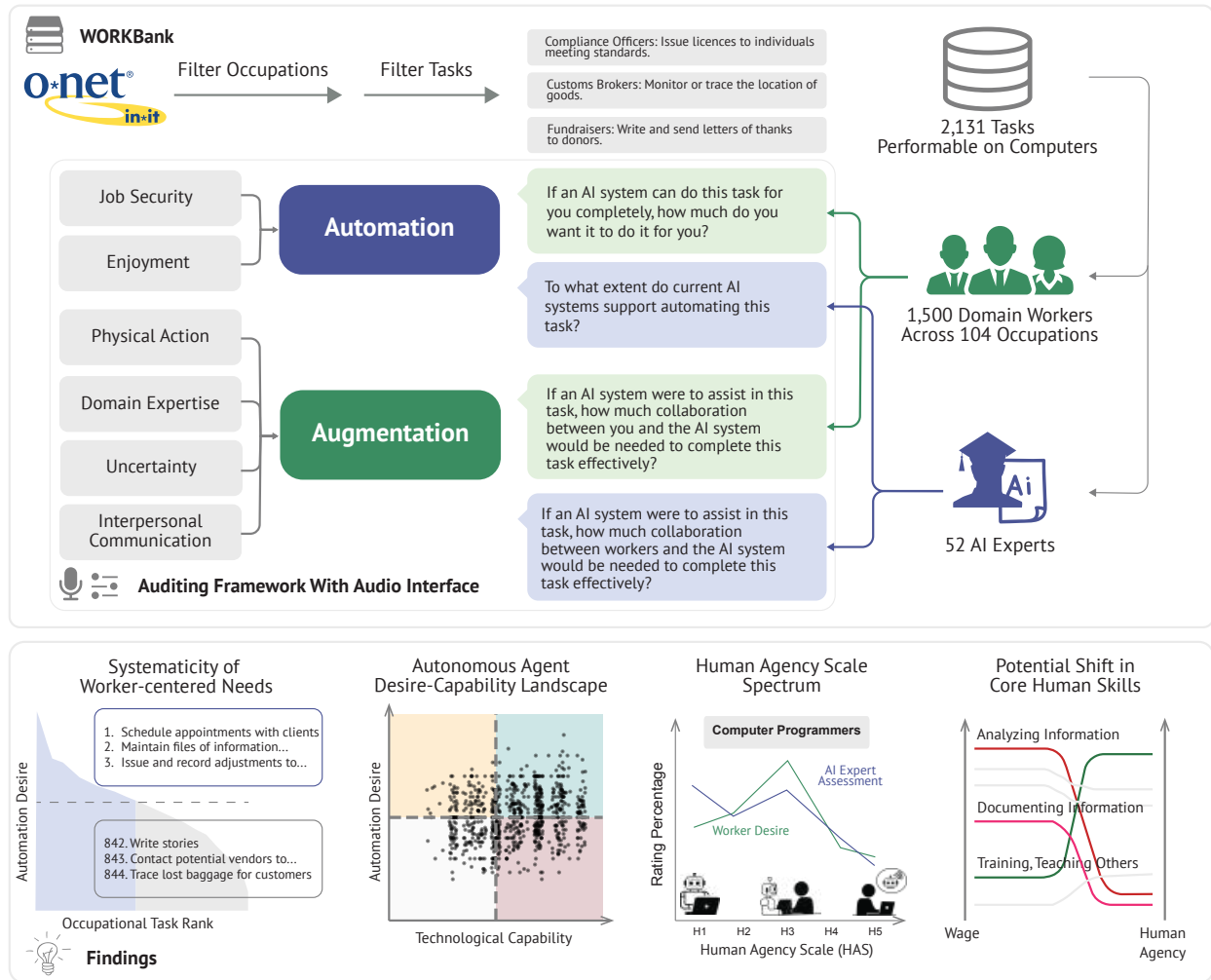- H2: AI agent needs minimal human input for optimal performance.

Figure 1: **Overview of the auditing framework and key insights.** The framework captures dual perspectives on automation and augmentation by eliciting both worker desires and expert assessments of technological capabilities. It guides participant reasoning through structured prompts and an audio-enhanced interface. We instantiate this framework to build the WORKBank database, enabling a data-driven analysis of worker-centered needs, the desire–capability landscape, the Human Agency Scale (HAS) spectrum, and implications for core human skills.

- H3: AI agent and human form equal partnership, outperforming either alone.
- H4: AI agent requires human input to successfully complete the task.
- H5: AI agent cannot function without continuous human involvement.

Unlike SAE driving automation levels (Committee et al., 2014) that adopt an "AI-first" perspective, HAS provides a human-centered lens for assessing both task properties and appropriate agent development approaches. Importantly, higher HAS levels are not inherently better—different levels suit different AI roles. Tasks at H1-H2 favor automation approaches, while H3-H5 tasks benefit from augmentation strategies. Understanding the ideal level of human involvement is essential both for workers seeking to adapt their skills and for developers aiming to build context-appropriate AI agents. For instance,
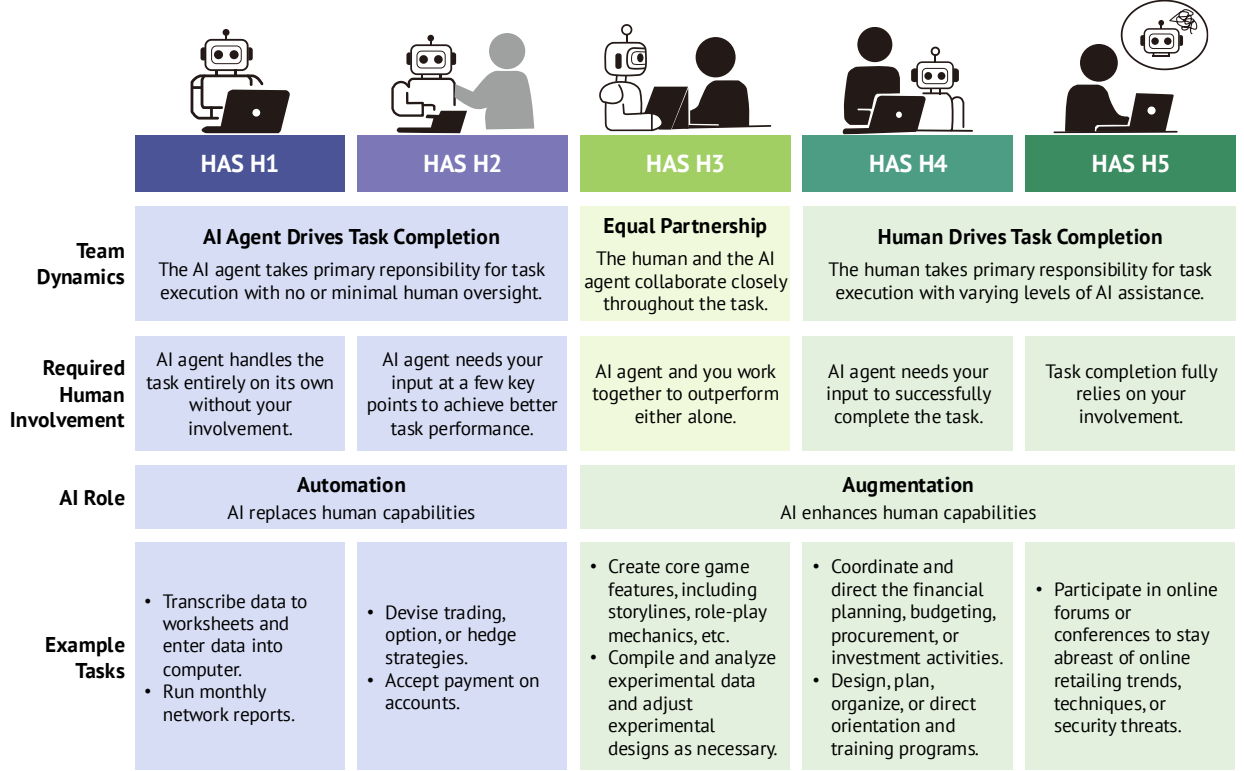
	HAS H1	HAS H2	HAS H3	HAS H4	HAS H5
Team Dynamics	**AI Agent Drives Task Completion** The AI agent takes primary reponsibility for task execution with no or minimal human oversight.		**Equal Partnership** The human and the AI agent collaborate closely throughout the task.	**Human Drives Task Completion** The human takes primary responsibility for task execution with varying levels of AI assistance.	
Required Human Involvement	AI agent handles the task entirely on its own without your involvement.	AI agent needs your input at a few key points to achieve better task performance.	AI agent and you work together to outperform either alone.	AI agent needs your input to successfully complete the task.	Task completion fully relies on your involvement.
AI Role	**Automation** AI replaces human capabilities		**Augmentation** AI enhances human capabilities		
Example Tasks	• Transcribe data to worksheets and enter data into computer. • Run monthly network reports.	• Devise trading, option, or hedge strategies. • Accept payment on accounts.	• Create core game features, including storylines, role-play mechanics, etc. • Compile and analyze experimental data and adjust experimental designs as necessary.	• Coordinate and direct the financial planning, budgeting, procurement, or investment activities. • Design, plan, organize, or direct orientation and training programs.	• Participate in online forums or conferences to stay abreast of online retailing trends, techniques, or security threats.

Figure 2: **Levels of Human Agency Scale (HAS).** We introduce the Human Agency Scale (*i.e.,* H1-H5) to quantify the team dynamics and degree of human involvement required. HAS provides a shared language to quantify automation vs. augmentation, complementing the traditionally "AI-first" perspective used in defining levels of automation. Importantly, higher HAS levels are not inherently better—different levels suit different AI roles.

fully autonomous agents shall be developed for H1 scenarios, while those agents for H3 scenarios must support meaningful coordination and communication with human collaborators (Shao et al., 2024b).

This five-level human agency scale (H1-H5) helps categorize tasks where AI is more suitable for automation (H1-H2) versus augmentation (H3-H5), where human agency remains critical.

2.3 Constructing A Worker-Centric Auditing Framework

Our auditing framework centers on the needs of workers. To support domain workers in providing well-calibrated feedback, we enable them to share their experiences and explain their thought processes using an audio-supported survey system. Concretely, for each task t, we first collect worker ratings on *automation desire* $A_w(t)$ and *desired HAS level* $H_w(t)$ using a 5-point Likert scale.

> **Likert Question for Collecting Automation Desire**
>
> If an AI system can do this task for you completely, how much do you want it to do it for you?
> 1: Not at all; 2: Slightly; 3: Moderately; 4: A lot; 5: Entirely

To support thoughtful ratings, we also scaffold worker responses through three key designs (survey details in Appendix A):

- **Audio-enhanced Reflection**: The survey begins with an audio-enhanced mini-interview exploring participants' occupational work and AI perspectives. This spoken format enables more natural reflection and helps workers more efficiently contextualize their ratings within their actual work experience, compared to slowly typing their experiences.

- **Quality Control via Task Familiarity Filtering**: Workers receive only occupation-relevant tasks and must confirm task familiarity before rating, ensuring assessments are grounded in their real experience rather than speculation.

- **Guided Consideration**: Before rating both automation desire $A_w(t)$ and desired HAS level $H_w(t)$, participants consider factors identified in prior literature—enjoyment and job security concerns for automation desire (Armstrong et al., 2024, Gödöllei and Beck, 2023), and task characteristics like physical actions, domain expertise requirements, uncertainty, and interpersonal elements for HAS preferences (Frank et al., 2019, Parasuraman, 2000, Shah and White, 2024).

2.4 Ensuring Dual Perspectives from Both Domain Workers and AI Experts

While worker perspectives provide invaluable insights into the social demand and acceptance of AI agents, they represent only one side of the integration equation. Domain workers, despite their deep task expertise, may have limited exposure to current AI capabilities and constraints. Thus, we complement workers' perspectives with expert assessments of *current automation capability* $A_e(t)$ and *feasible HAS level* $H_e(t)$ from AI researchers and practitioners. This dual perspective reveals the readiness for AI agent integration and allows us to identify alignment or gaps between worker desires and technological feasibility.

Concretely, these experts assess $A_e(t)$ and $H_e(t)$ using the same rubrics, drawing on their understanding of existing systems' strengths and limitations. Contrasting $A_w(t)$, $H_w(t)$ and $A_e(t)$, $H_e(t)$ enables us to understand what require future breakthroughs, identify alignments and misalignments between worker preferences and technological development, and inform development priorities .

2.5 Instantiating the Audit Framework to Derive WORKBank

Taking into account these aforementioned design principles together, we then apply our auditing framework to develop **W**orker **O**utlook & **R**eadiness **K**nowledge Bank (WORKBank). Concretely, we source computer-compatible tasks performed at least monthly from the U.S. Department of Labor's O*NET Database (details in Appendix C.1). These tasks reflect complex, multi-step workflows central to our focus. For example, the task "Credit Analysts: Analyze credit data and financial statements to determine the degree of risk involved in extending credit or lending money" entails data analysis, risk evaluation, and decision-making. After filtering, 2,131 tasks across 287 occupations remain.
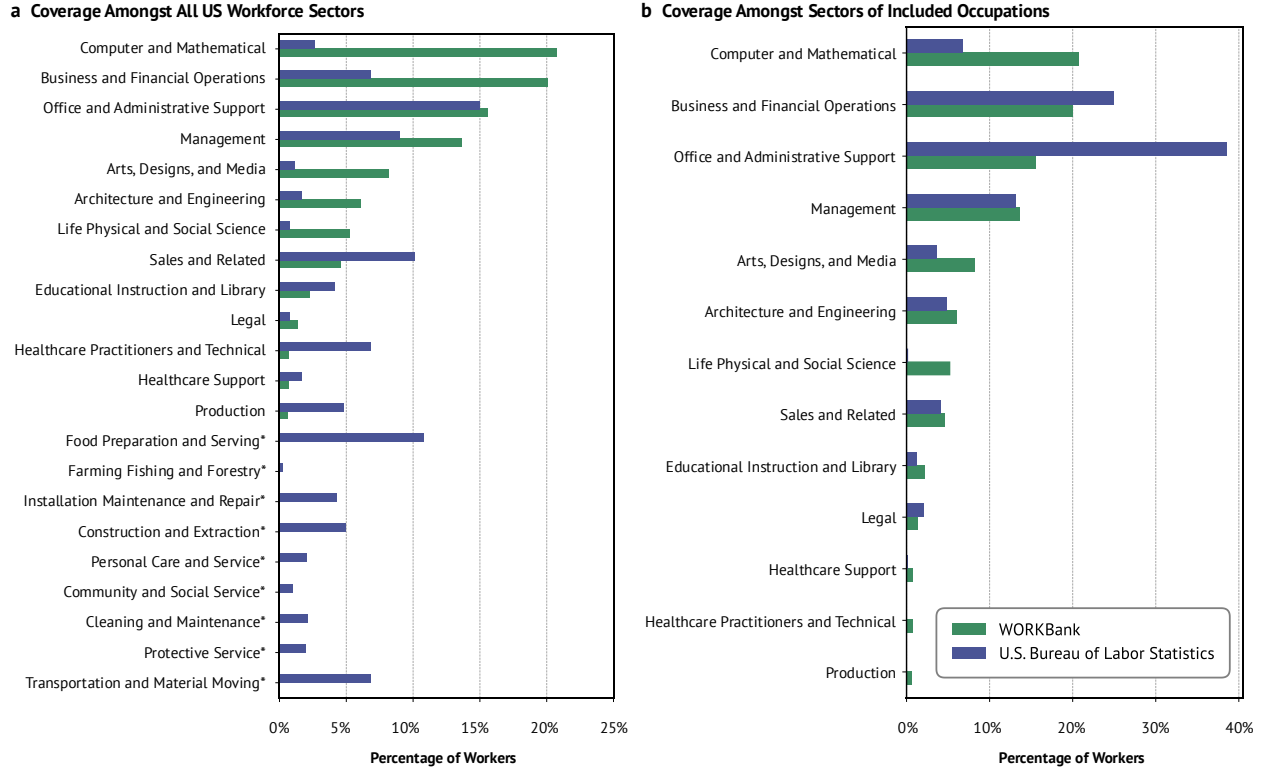
a Coverage Amongst All US Workforce Sectors

b Coverage Amongst Sectors of Included Occupations

Figure 3: **Sector-level distribution of workers in the WORKBank database compared to U.S. workforce statistics from the Bureau of Labor Statistics. a**, Comparison between WORKBank worker distribution and the U.S. workforce employment statistics across all sectors (sectors not included in WORKBank marked with an asterisk). **b**, Comparison between WORKBank worker distribution and the U.S. workforce employment statistics limited to the 104 occupations included in our database.

We developed an IRB-approved, self-hosted survey interface and distributed it through crowdsourcing platforms and targeted LinkedIn outreach. Between January and May 2025, recruitment through Prolific, Upwork, and LinkedIn yielded 1,676 participants who provided 7,016 task ratings. After filtering for adequate representation (≥ 10 participants per occupation), we obtained assessments from 1,500 individuals across 104 occupations and calculated average worker ratings $A_w(t)$ and $H_w(t)$ for each task. We evaluate the representativeness of WORKBank by comparing its sector-level distribution with U.S. workforce data from the Bureau of Labor Statistics (Appendix D.2). As shown in Figure 3, the comparisons suggest that our database captures a broad and representative cross-section of the U.S. workforce at the sector level.

For expert assessments, we recruited 52 AI experts—PhD researchers and industry practitioners with experience in AI agent R&D. Each task was independently assessed by at least two experts, with additional reviews ensuring rating consistency (standard deviation ≤ 1). Inter-annotator agreement, measured by Krippendorff's α, was 0.539 for $A_e(t)$ and 0.511 for $H_e(t)$ (see robustness analysis details in Appendix B).

Combining these complementary data sources, we construct **W**orker **O**utlook & **R**eadiness **K**nowledge Bank (WORKBank), the first database to capture both worker desires and AI agents' technological capabilities for occupational tasks.

7

3 Results

Leveraging the rich data within WORKBank, we examine where workers most desire automation by AI agents and where they resist it, whether current AI capabilities and R&D align with these preferences, what opportunities exist for AI to augment rather than replace human labor, and how the presence of AI agents might reshape the demand for human skills.

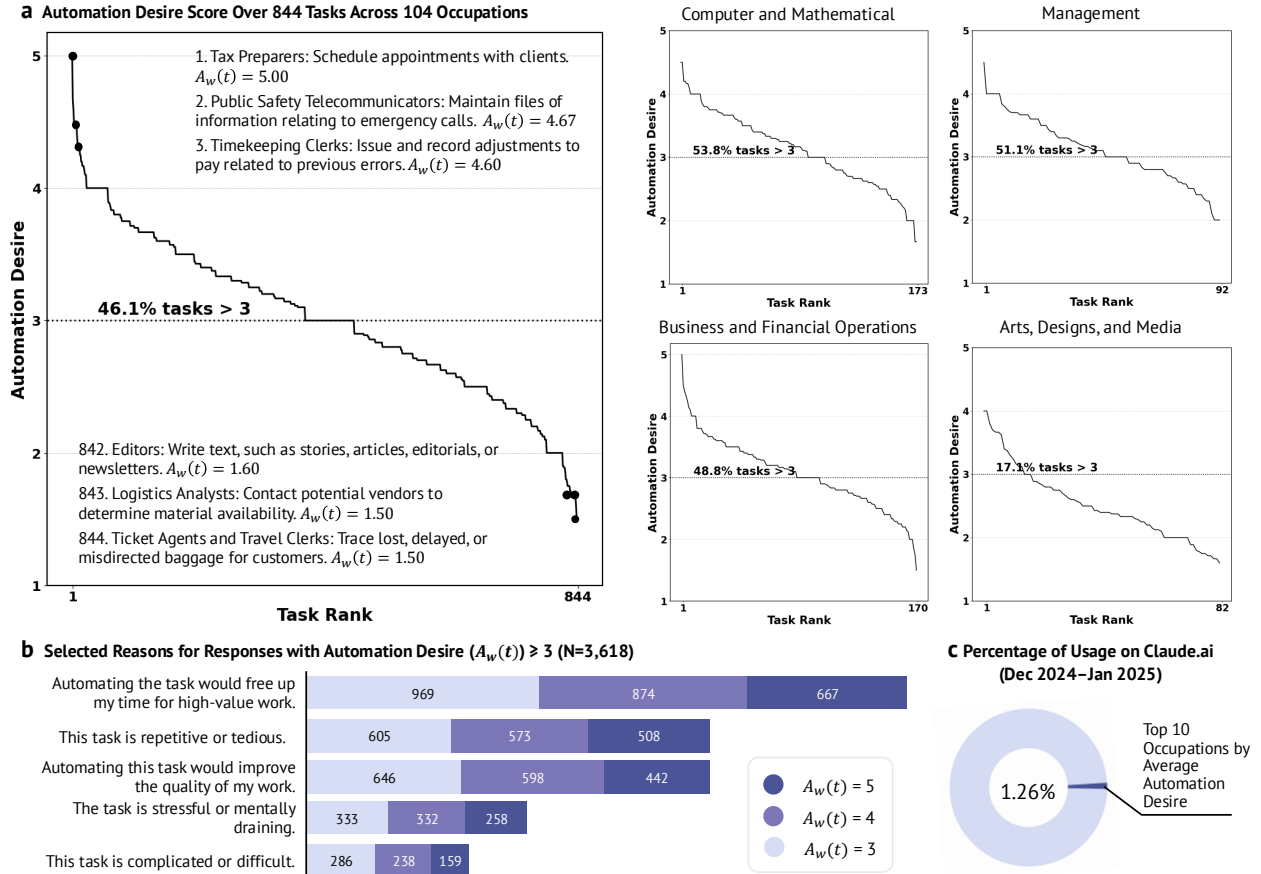## 3.1 Worker-centered Views on Occupational Task Automation

Figure 4: **First-hand data from domain workers reveals positive attitudes towards AI agent automation on certain occupational tasks, particularly due to perceived benefits such as freeing up time for high-value work. However, the sentiment varies notably across sectors. a**, Automation desire scores $A_w(t)$ over 844 occupational tasks, ranked based on WORKBank data, together with sector-specific breakdowns. The distribution indicates a mixed attitude, revealing high diversity of needs and preferences of workers that should be considered in AI agent R&D. **b**, Reported reasons for responses with $A_w(t) \geq 3$. The most selected reason—"Automating the task would free up my time for high-value work"—accounts for 69.38% of the responses. **c**, Comparison with usage data from Claude.ai, a LLM-based chatbot (Dec 2024-Jan 2025, from Handa et al. (2025)), shows that the top 10 occupations with the highest average automation desire represent only 1.26% of total usage. This highlights the importance of directly soliciting worker input, as usage data may lag behind actual workplace needs.

Where do workers desire AI agent automation? We first examine domain workers' attitudes toward automating their occupational tasks. In Figure 4 **a**, we rank tasks by their average worker automation desire scores $A_w(t)$. We find that for 46.1% of tasks, workers currently performing them express a positive attitude (*i.e.*, $A_w(t) > 3$) toward AI agent automation, even after explicitly considering concerns such as job loss and reduced enjoyment, as guided by our auditing framework. On the other hand, the distribution indicates a mixed attitude, with 7.11% tasks receiving $A_w(t) \geq 4$ and 6.16% receiving $A_w(t) \leq 2$. To better understand these preferences, Figure 4 **b** aggregates selected reasons given for pro-automation responses ($A_w(t) \geq 3$). The most cited motivation—"freeing up time for high-value work"—was selected in 69.38% of cases. Other common reasons include task repetitiveness (46.6%), stressfulness (25.5%), and opportunities for quality improvement (46.6%). The overall pattern suggests that AI agents could play a supportive role, enabling workers to offload low-value or burdensome tasks, rather than serving as replacements in a zero-sum dynamic.

Does existing LLM usage reflect worker desires? Notably, when we compare our findings with usage data from Claude.ai, an LLM-based chatbot used between Dec 2024 and Jan 2025 (Handa et al., 2025), we find that the top 10 occupations with the highest average automation desire account for only 1.26% of total usage. This mismatch highlights a disconnect: occupations where workers most desire automation are currently underrepresented in LLM usage. This suggests that existing usage patterns may be skewed toward early adopters or specific job types, rather than reflecting broader demand. Such a gap reinforces the value of our worker-centric audit, which surfaces latent needs that may not yet appear in usage logs.

Where do workers resist AI agent automation? We analyze audio response data using LLM-based topic modeling to identify the primary concerns workers expressed regarding the use of AI agents in their work (see Appendix F.1). Among our survey participants, 28.0% expressed fears, concerns, or negative sentiment when answering the question "How do you envision using AI in your daily work?". Among these workers, the three most prominent concerns identified are: (1) lack of trust in AI systems' accuracy, capability, or reliability (45.0%), (2) fear of job replacement (23.0%), and (3) the absence of human qualities or capabilities in AI (16.3%).

When discussing the absence of human qualities, workers express specific concerns about losing a "human touch" in their work, diminishing creative control, and the desire to maintain agency in decision-making. This sentiment echoes our quantitative findings from the breakdown of automation desire scores across sectors (Figure 4 **a**). In these sector-level breakdowns, the "Arts, Designs, and Media" sector stands out, with only 17.1% of tasks receiving positive desire ratings (> 3 on a 5-point Likert scale). Audio responses from participants in this sector reveal nuanced opposition to automating content creation, such as: *"I want it to be used for seamlessly maximizing workflow and, you know, making things less repetitive and tedious and arduous with workflow. No content creation," "I would never use AI to like replace artists. I would be more for personal [project management] use, if anything," "AI can be a game-changer in data architect workflow, helping to improve efficiency, accuracy and even creativity. But I create my design by myself. For research, I use AI"*.

3.2 Desire-Capability Landscape for AI Agents in the Workplace

Contrasting worker and AI expert perspectives delineate four task zones. While workers' preferences offer valuable guidance for socially beneficial AI agent deployment, delivering impact ultimately depends on aligning those preferences with technical feasibility. To investigate this, we jointly consider the worker-rated automation desire $A_w(t)$ and expert-assessed technological capability $A_e(t)$, visualized as a desire-capability landscape in Figure 5 **a**. This landscape divides into four zones:
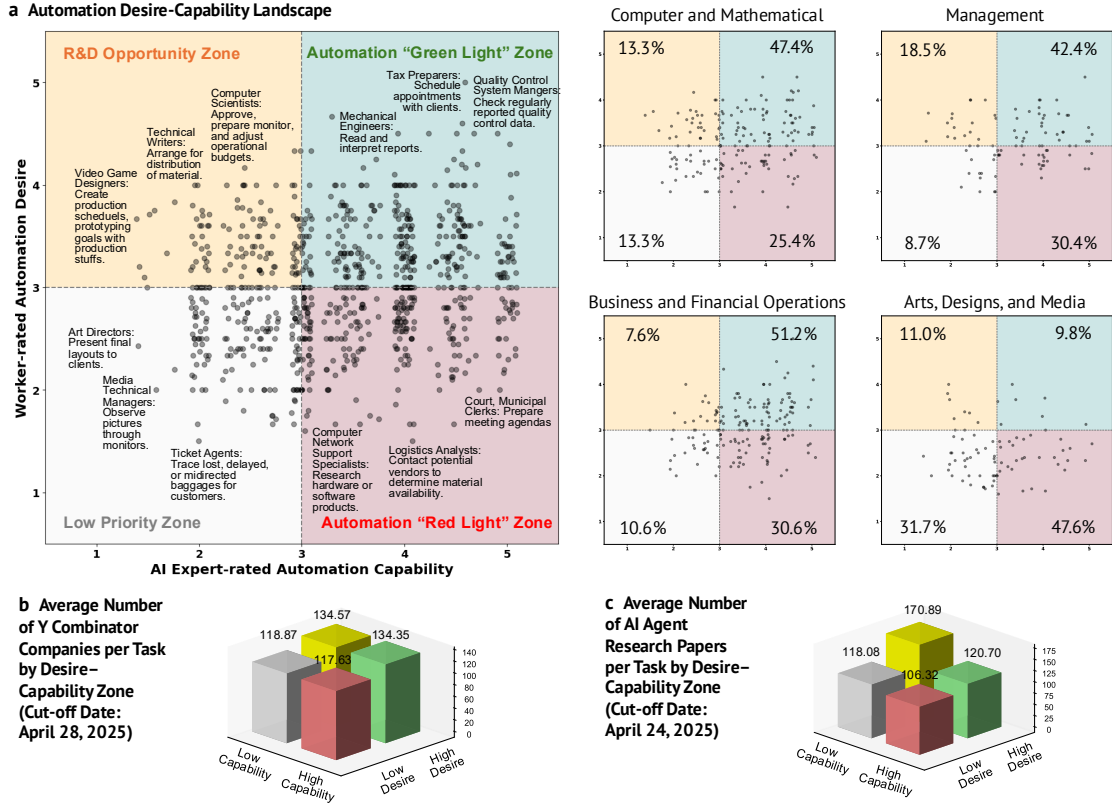
Figure 5: **Integrating worker and AI expert perspectives divides the automation landscape into four zones: Automation "Green Light" Zone, Automation "Red Light" Zone, R&D Opportunity Zone, and Low Priority Zone. a**, Tasks from WORKBank are plotted in this desire-capability landscape. **b**, We collect Y Combinator (YC) companies and map them to tasks based on the description on their official YC detail pages using `gpt-4.1-mini`. The average number of YC companies per task shows no significant difference across zones, highlighting the importance of steering more investment toward the Automation "Green Light" Zone and R&D Opportunity Zone. **c**, We collect AI agent research papers from arXiv and evaluate their applicability to each occupational task in our database using `gpt-4.1-mini`. Encouragingly, the paper-task mappings are concentrated more in the R&D Opportunity Zone, though increased emphasis on this area remains desirable.

1. **Automation "Green Light" Zone**: Tasks with both high automation desire and high capability. These are prime candidates for AI agent deployment with the potential for broad productivity and societal gains.
2. **Automation "Red Light" Zone**: Tasks with high capability but low desire. Deployment here warrants caution, as it may face worker resistance or pose broader negative societal implications.
3. **R&D Opportunity Zone**: Tasks with high desire but currently low capability. These represent promising directions for AI research and development.
4. **Low Priority Zone**: Tasks with both low desire and low capability. These are less urgent for AI agent development.

Tasks from WORKBank are broadly distributed across the landscape, with no strong correlation between $A_w(t)$ and $A_e(t)$ (Spearman $\rho = 0.17$, $p < 1e-6$). Overall, automation desire shows a negative

correlation with both job loss concern (Spearman $\rho = -0.223, p < 5e-11$) and enjoyment (Spearman $\rho = -0.284, p < 3e-17$). This suggests both alignments and misalignments between worker desires and technological capabilities, with a consistent pattern: workers are less inclined to have AI agents automate tasks they enjoy or feel vulnerable about potential job loss, consistent with findings from prior literature (Armstrong et al., 2024, Gödöllei and Beck, 2023).

Mapping investment to the desire-capability landscape reveals critical mismatches. To better understand where current investments are concentrated, we used Y Combinator (YC) companies[1] as a proxy and mapped them to the tasks in our database using gpt-4.1-mini (one company could be mapped to multiple tasks, see Appendix C.4 for details). As shown in Figure 5 **b**, the company-task mappings are relatively evenly spread across the four zones. Most mapped tasks are concentrated in occupations related to software development and business analysis, with the top five occupations being: Computer and Information Systems Managers, Computer Programmers, Computer Systems Engineers/Architects, Software Quality Assurance Analysts and Testers, and Business Intelligence Analysts. 41.0% of YC companies are mapped to Low Priority and Automation "Red Light" Zone; while many promising tasks within the "Green Light" Zone and Opportunity Zone remain under-addressed by current investments.

AI agent research papers show an emphasis on the R&D Opportunity Zone, but remain concentrated on a limited set of tasks. Following a similar methodology, we gathered research papers related to AI agents from arXiv[2] and analyzed their alignment with various tasks to determine the distribution of research efforts. Figure 5 **c** shows that research papers are more concentrated in the R&D Opportunity Zone. While encouraging, the focus remains largely confined to computer science and engineering domains. The top three tasks covered are: (1) Computer and Information Research Scientists: Apply theoretical expertise and innovation to create or apply new technology, such as adapting principles for applying computers to new uses (1,169 papers); (2) Computer and Information Research Scientists: Analyze problems to develop solutions involving computer hardware and software (1,132 papers); (3) Computer Programmers: Perform or direct revision, repair, or expansion of existing programs to increase operating efficiency or adapt to new requirements (1,109 papers). These findings highlight the need to expand research efforts beyond a few domains to better support tasks in the R&D Opportunity Zone, ensuring that future AI agents address a wider range of high-desire, high-impact opportunities.

3.3 Human Agency Scale (HAS) Spectrum

Beyond automation, AI agents hold promise for augmenting human work. To understand where and how this augmentation may occur, we analyze the distribution of both worker-desired HAS levels ($H_w(t)$) and expert-assessed feasible HAS levels ($H_e(t)$) across tasks within each occupation.

Where do worker desires and expert assessments diverge most on the Human Agency Scale? Each task in WORKBank is assigned a worker-desired and expert-rated HAS level via majority vote. Among 844 tasks, 26.9% receive matching levels from workers and AI experts. Figure 6 **a** shows workers generally prefer higher levels of human agency than what experts deem technologically necessary, with 47.5% of tasks fall into the lower triangle of the matrix. To quantify this divergence, we compute the Jensen-Shannon Distance (JSD) between the distributions of $H_w(t)$ and $H_e(t)$ at the occupation level. Disagreements are most pronounced in the lower HAS range as Table 5 shows that five of the ten occupations with the highest JSD scores are also those that experts rate as H1 dominant.This

---

[1] https://www.ycombinator.com/companies
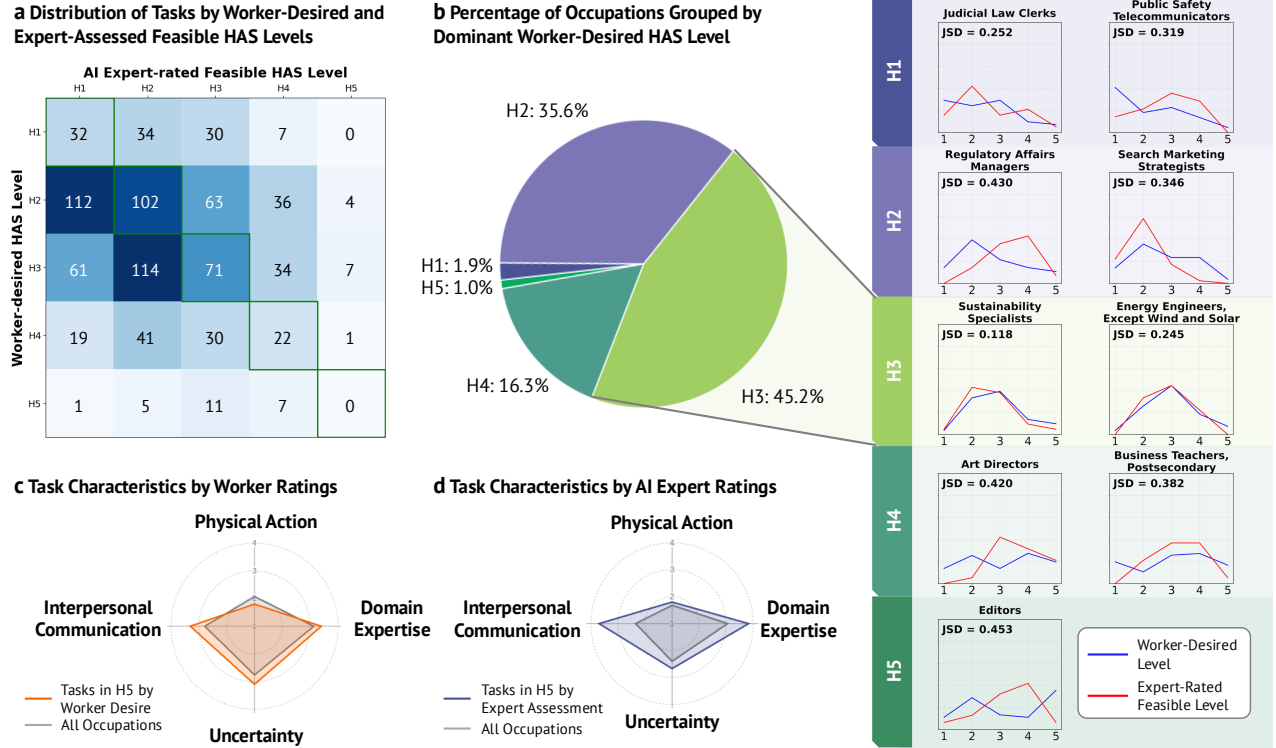[2] http://arxiv.org

Figure 6: **Distributions on the Human Agency Scale (HAS) reveal diverse patterns of AI agent integration across occupations and underscore opportunities for human–agent collaboration. a**, Comparison between worker-desired HAS levels ($H_w(t)$) and expert-assessed feasible HAS levels ($H_e(t)$) shows that workers generally prefer higher levels of human agency than what experts deem technologically necessary. **b**, Distribution of occupations by their dominant worker-desired HAS level shows that most occupations cluster around H3 and helps identify occupations that are at the poles of the agency spectrum. Each subplot displays the task-level distributions of worker-desired and expert-assessed HAS levels for a given occupation. Jensen-Shannon divergence (JSD) quantifies the difference between these distributions (see top 10 in Table 5). Full occupation-level results are shown in Figure 10. **c**, Radar plot of task characteristics based on worker ratings indicates that tasks in the H5 region are particularly associated with Interpersonal Communication. **d** Radar plot based on expert ratings shows that tasks in H5 are marked by strong Interpersonal Communication and Domain Expertise components.

signals potential frictions as AI adoption progresses. Ensuring socially responsible deployment of AI agents and supporting workers currently performing low-HAS tasks warrant further scrutiny.

What are the common patterns across the Human Agency Scale spectrum? As illustrated in Figure 6 **b** (with full results in Figure 10), many occupations (*e.g.*, "Sustainability Specialists", "Energy Engineers") exhibit an inverted-U shaped distribution for both $H_w(t)$ and $H_e(t)$. While this trend in expert assessments might reflect current technological limitations—*i.e.*, AI agents are not yet capable of fully replacing human involvement in most tasks—it is notable that workers in many domains also prefer a balanced, collaborative partnership with AI. H3 emerges as the dominant worker-desired level in 47 out of 104 occupations analyzed.

Which occupations stand out on the Human Agency Scale? Beyond the general inverted-U trend, we examine occupations at the extremes of the HAS spectrum. Lower HAS levels signify areas of greater potential AI exposure. According to AI experts' ratings, 16 out of 104 occupations are predominantly H1, even based on current capability estimates. These include roles such as "Computer Programmers", "Proofreaders and Copy Makers", and "Travel Agents". Occupations within the same sector also exhibit distinct trends in HAS levels. For example, "Computer Programmers" and "Information Technology Project Managers" display markedly different distributions (H1 vs. H4) when assessed by AI experts. Compared to Eloundou et al. (2023), which provides an early analysis of LLMs' labor market impact and finds higher-wage occupations to be more exposed, our results show that while most occupations in WORKBank are indeed exposed to AI agents and do not fall into H5 (essential human involvement), those involving more routine tasks and easily verifiable outcomes tend to require lower human agency.

Very few occupations are characterized by a dominant HAS Level 5 (indicating essential human involvement). "Editors" is the only occupation where workers predominantly desire H5. According to AI expert assessments, only "Mathematicians" and "Aerospace Engineers" fall into this category. Representative work descriptions from worker transcripts for these occupations are provided in Appendix F.2. We further investigate what distinguishes H5 tasks. Among the four quantitative dimensions in our auditing framework (Figure 1), worker ratings highlight Interpersonal Communication as a defining feature of H5 tasks (Figure 6 **c**), while expert ratings emphasize both Interpersonal Communication and Domain Expertise (Figure 6 **d**).

What forms of human-agent collaboration do workers envision? Figure 6 **b** suggests strong potential for collaborative AI. Our analysis of audio transcripts supports this: the vast majority of workers express either a desire or openness to collaborating with AI to enhance their work. Analyzing these narratives further illustrates how workers concretely envision human–agent partnerships. The most common paradigm is "role-based" AI support (described by 23.1% of workers), where individuals anticipate *utilizing AI systems that embody specific roles or personalized functions ("[I would like to have an AI agent] trained to automatically analyze the quality control reports of raw sequencing data (e.g., FastQC output) and flag potential issues with specific samples or sequencing runs, [...] and suggesting appropriate preprocessing steps." "It's just for me to set up my AI.")*. Furthermore, 23.0% of workers express a desire for AI systems to function as a supportive assistant for some or all aspects of their workflow (*"[I envision the AI agent] as an assistant who is doing research for me. However, I review every answer because we cannot rely on its accuracy."*), while 16.5% mention pure automation by AI for some aspects of their workflow.

3.4 The Potential Shift of Core Human Skills

Variation in Human Agency Scale (HAS) across occupations suggests that certain types of human work are more susceptible to automation, while others hold greater potential for augmentation. To this end, we examine the characteristics of tasks that require high human agency to understand how AI agents may shift skill demands.

Translating task-level changes into implications for education and skill training has long been a key lens for analyzing technological transformation, notably pioneered by Autor et al. (2003) in the wave of computers. To operationalize this lens, we align each task with its related skills (Generalized Work Activities) as defined by O*NET (Appendix E.6). For example, the task "Financial Managers: Approve, reject, or coordinate the approval or rejection of lines of credit or commercial, real estate, or personal loans" will be mapped to "making decisions and solving problems" and "guiding, directing, and motivating subordinates". We compute the average expert-assessed human agency level $H_e(t)$ for each skill to estimate the degree of human involvement required as AI agents enter the workplace. We also com-
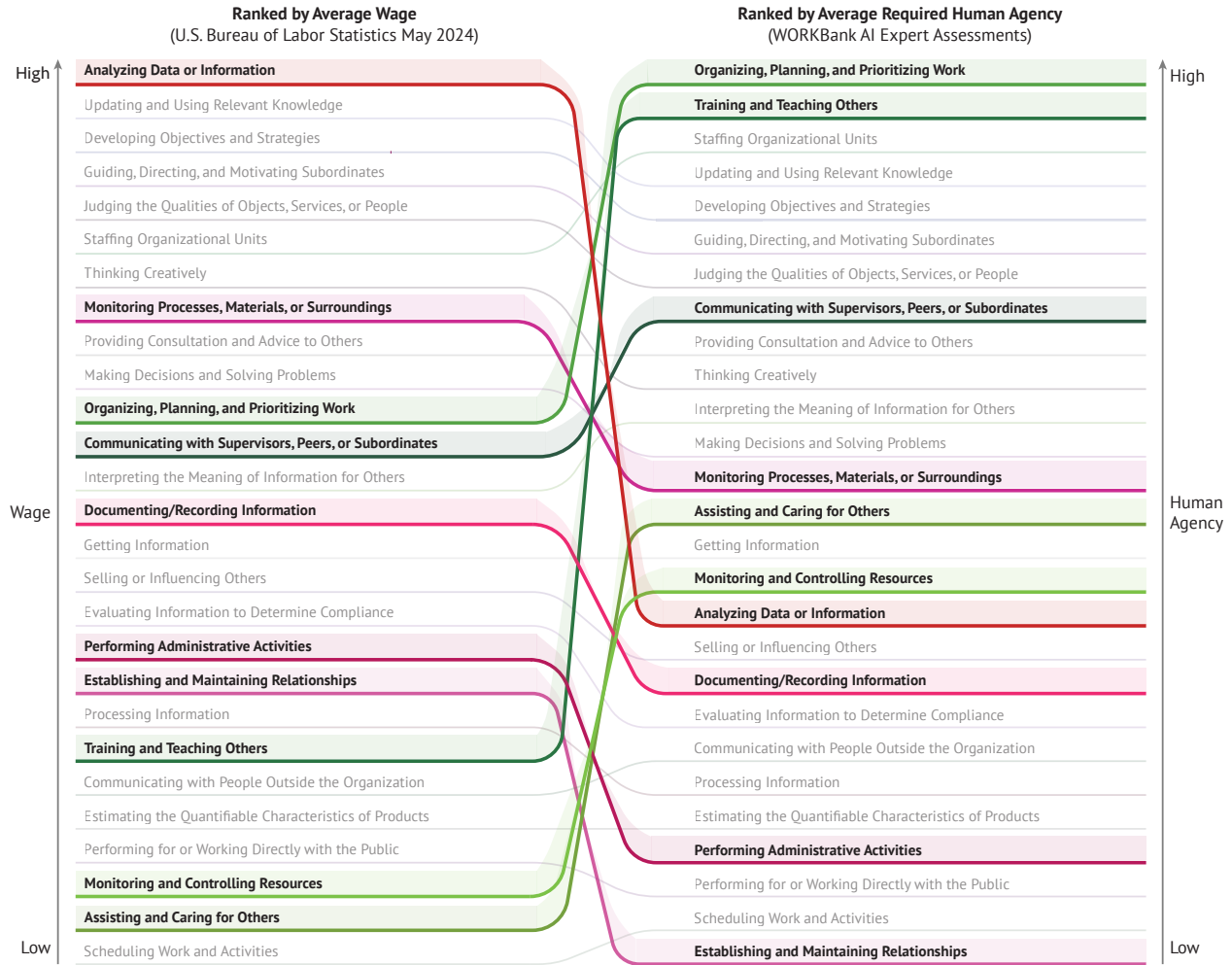
Figure 7: **Comparing skill rankings by average wage and required human agency.** Each line represents a skill (Generalized Work Activity) mapped from O*NET tasks. Based on the skill-task mapping, we compute the average wage using data from the U.S. Bureau of Labor Statistics (May 2024) and the average expert-assessed human agency level $H_e(t)$ to indicate the degree of human involvement required as AI agents enter the workplace. Skills are ranked by average wage (left) and average required human agency (right). The figure highlights the top five skills with the largest upward (green) and downward (red) shifts in rank, suggesting a potential shift in valued workplace skills—from information processing toward interpersonal and organizational competencies. See Appendix E.6 for skill analysis details and full skill descriptions.

pute the average wage for each skill, using data from the U.S. Bureau of Labor Statistics (Appendix C.2). Wage serves as a proxy for the current economic value of each skill (Dey and Loewenstein, 2019).

As shown in Figure 7, by comparing skill rankings by average wage and required human agency, our analysis reveals three emerging trends:

1. **Shrinking demand for information-processing skills.** Skills related to analyzing data and updating knowledge—while common in today's high-wage occupations (as shown in the left side of Figure 7 in red color)—are less prominent in tasks that demand high human agency.

2. **Greater emphasis on interpersonal and organizational skills.** Skills involving human interaction, coordination, and resource monitoring are more frequently associated with high-HAS tasks (as shown in the left side of Figure 7 in green color), even if they are not currently prioritized in wage-based evaluations.

3. **High-agency skills span diverse aspects.** The top 10 skills with the highest average required human agency encompass a broad range, from interpersonal and organizational abilities to decision-making and quality judgment.

These findings provide early signals of how AI agent integration may reshape core occupational competencies. As workplace AI agents continue to evolve, longitudinal tracking of task-level changes could yield further insights into how human roles and required skills evolve.

4 Related Work

Digital AI Agents   The aspiration to build AI agents capable of dynamically directing their own processes to accomplish complex goals dates back to the early days of artificial intelligence (Genesereth and Nilsson, 1987, McCarthy, 1959). Recent advances in foundational models, particularly large language models (LLMs), have sparked a surge in the development of digital AI agents that leverage LLMs to plan actions and interface with external tools (Sumers et al., 2023, Wang et al., 2024a). These agents have demonstrated the ability to carry out workflows across diverse domains, including software engineering (Wang et al., 2024b, Yang et al., 2024), analytical writing (Jiang et al., 2024, Shao et al., 2024a), and customer support (Yao et al., 2024).

While many of these agents are designed for full automation, they can also be structured to collaborate with humans. Collaborative Gym (Shao et al., 2024b) pioneered the concept of human-agent collaboration, demonstrating that for certain tasks, joint human-agent performance can surpass that of fully autonomous agents, even when those agents are capable of completing the tasks independently. This underscores the potential of AI agents to augment, rather than simply replace, human labor (Brynjolfsson, 2022). The auditing framework proposed in this work systematically examines augmentation versus automation by introducing the Human Agency Scale (HAS), which evaluates the level of ideal human involvement across different workflows.

One limitation of prior work on AI agents is its frequent focus on a narrow set of domains. Existing benchmarks, such as GAIA (Mialon et al., 2023), AgentBench (Liu et al., 2023), OSWorld (Xie et al., 2024), while valuable for assessing agent capabilities, often rely on task collections that are curated in a constrained manner. Such approach, while useful for capability evaluation, fails to provide a holistic and worker-centric understanding of how these agents could be integrated into the broader workforce. By sourcing tasks from the U.S. Department of Labor's O*NET database, our work provides a more comprehensive and systematic understanding of the potential landscape for digital AI agents.

The Economic Impacts of Generative AI   A broad body of work in digital economics has examined the implications of AI, spanning from early machine learning models (Brynjolfsson and Mitchell, 2017) and computer vision systems (Svanberg et al., 2024) to the recent surge of large language models (LLMs) and generative AI (Demirci et al., 2025, Eloundou et al., 2023, Handa et al., 2025, Hoffmann et al., 2024). Following the launch of ChatGPT, Eloundou et al. (2023) provided an early analysis of LLMs' potential labor market impact, estimating that approximately 80% of the U.S. workforce has at least some tasks exposed to LLM capabilities. However, their analysis did not incorporate the dimension of

worker desire and focused primarily on LLM via ChatGPT or the OpenAI playground rather than the broader scope of AI agents. More recent work leveraging real user data from Claude.ai, a state-of-the-art LLM chatbot, to identify which economic tasks users actually perform with AI (Handa et al., 2025). In parallel, field studies in customer support organizations have shown that AI-assisted chatbots can improve worker productivity (Brynjolfsson et al., 2025). As AI agents continue to evolve beyond standalone LLM chatbots, our study provides an early audit of their readiness for workplace integration.

5   Conclusion

Advancements in AI agents are unlocking a wide range of possibilities that may fundamentally reshape the workplace. This paper presents the first large-scale audit of both worker desire and technological capability for AI agents in the context of automation and augmentation. Based on data collected between January and May 2025, we construct the WORKBank database and find that domain workers generally express positive attitudes toward AI agent automation, particularly for repetitive and low-value tasks.

By integrating both worker and expert perspectives, we introduce the automation desire–capability landscape, which offers actionable insights for prioritizing AI agent research and investment. Besides the traditional automate-or-not dichotomy, our Human Agency Scale (HAS) uncovers diverse patterns of AI integration across occupations, with a dominant inverted-U trend that underscores the potential for human–agent collaboration.

Beyond informing AI agent research and deployment strategies, our findings also have implications for workforce development. As AI agents reshape the demand for core human skills, our findings suggest that examining strategies for worker reskilling and retraining is a valuable direction for future research.

Limitations   While our audit offers a comprehensive snapshot of worker perspectives and technological capabilities of workplace AI agents, several limitations should be considered:

First, our quantitative assessments are grounded in existing occupational tasks defined by the O*NET database, which does not account for new tasks that may emerge as AI agents become more integrated into the workplace. Further analysis of the open-ended worker transcripts could uncover emerging task patterns and enrich our understanding of evolving occupational tasks.

Second, although we guided participants to reflect on potential job loss and task enjoyment, domain workers may still lack full awareness of the evolving capabilities and limitations of AI agents (Hazra et al., 2025), potentially shaping their responses. We partially mitigate this limitation by including only occupations with at least 10 worker responses in the AI Agent WORKBank. Robustness checks and further discussion are provided in Appendix B, and complement it with AI experts' assessment.

From an incentive perspective, some workers may also withhold honest feedback due to concerns about job security or surveillance. We recognize that this is a real concern, which is why we're committed to a worker-focused approach that surfaces real concerns and co-designs systems that reflect workers' values. We see such resistance as a critical input that helps guide responsible deployment. By prioritizing workers' perspectives, we enable workers to play an active role in shaping the future of work, rather than just adapting to it.

Third, the current version of AI Agent WORKBank only covers 104 occupations, a subset of the 287 computer-using occupations identified with the O*NET database. In our study, we launched the survey interface on Prolific, Upwork, and LinkedIn in January 2025 and concluded data collection in May 2025 to ensure temporal consistency. These 104 occupations were retained after filtering for adequate representation (≥ 10 participants per occupation). While our database exhibits strong

coverage and demographic representativeness (see Appendix D), our findings may not cover the full picture of AI agents for the workplace.

Finally, the AI Agent WORKBank reflects the present state of generative AI and agentic systems as of early 2025. As AI capabilities continue to evolve, the landscape of feasible and desirable agent-supported tasks will likely shift. While our framework offers a timely and structured baseline, future iterations of this audit will be essential for tracking long-term trends and informing the responsible development of workplace AI systems.

Acknowledgements

References

Ben Armstrong, Valerie K. Chen, Alex Cuellar, Alexandra Forsey-Smerek, and Julie A. Shah. Automation from the worker's perspective, 2024. URL https://arxiv.org/abs/2409.20387.

David H Autor. Why are there still so many jobs? the history and future of workplace automation. *Journal of economic perspectives*, 29(3):3–30, 2015.

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

Erik Brynjolfsson. The turing trap: The promise & peril of human-like artificial intelligence. *Daedalus*, 151(2):272–287, 2022.

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

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative ai at work. *The Quarterly Journal of Economics*, page qjae044, 2025.

Cristiano Castelfranchi. Guarantees for autonomy in cognitive agent architecture. In *International workshop on agent theories, architectures, and languages*, pages 56–70. Springer, 1994.

SAE On-Road Automated Vehicle Standards Committee et al. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standard J*, 3016:1, 2014.

Ozge Demirci, Jonas Hannane, and Xinrong Zhu. Who is ai replacing? the impact of generative ai on online freelancing platforms. *Management Science*, 2025.

Matthew S Dey and Mark A Loewenstein. *On job requirements, skill, and wages*. US Department of Labor, US Bureau of Labor Statistics, Office of Employment . . . , 2019.

Andrea L Eisfeldt, Gregor Schubert, Miao Ben Zhang, and Bledi Taska. The labor impact of generative ai on firm values. *Available at SSRN 4436627*, 2023.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. Gpts are gpts: An early look at the labor market impact potential of large language models, 2023. URL https://arxiv.org/abs/2303.10130.

Rosanna Fanni, Valerie Eveline Steinkogler, Giulia Zampedri, and Jo Pierson. Enhancing human agency through redress in artificial intelligence systems. *AI & society*, 38(2):537–547, 2023.

Morgan R Frank, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, et al. Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14): 6531–6539, 2019.

M. R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. 1987.

Anna F. Gödöllei and James W. Beck. Insecure or optimistic? employees' diverging appraisals of automation, and consequences for job attitudes. *Computers in Human Behavior Reports*, 12:100342, 2023. ISSN 2451-9588. doi: https://doi.org/10.1016/j.chbr.2023.100342. URL https://www.sciencedirect.com/science/article/pii/S2451958823000751.

Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, et al. Which economic tasks are performed with ai? evidence from millions of claude conversations. *arXiv preprint arXiv:2503.04761*, 2025.

Sanchaita Hazra, Bodhisattwa Prasad Majumder, and Tuhin Chakrabarty. Ai safety should prioritize the future of work. *arXiv preprint arXiv:2504.13959*, 2025.

Manuel Hoffmann, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu. Generative ai and the nature of work. Technical report, CESifo Working Paper, 2024.

Yucheng Jiang, Yijia Shao, Dekun Ma, Sina Semnani, and Monica Lam. Into the unknown unknowns: Engaged human learning through participation in language model agent conversations. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 9917–9955, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.554. URL https://aclanthology.org/2024.emnlp-main.554/.

Michelle S Lam, Janice Teoh, James A Landay, Jeffrey Heer, and Michael S Bernstein. Concept induction: Analyzing unstructured text with high-level concepts using lloom. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–28, 2024.

Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv:2308.03688*, 2023.

John McCarthy. Programs with common sense, 1959.

Grégoire Mialon, Clémentine Fourrier, Thomas Wolf, Yann LeCun, and Thomas Scialom. Gaia: a benchmark for general ai assistants. In *The Twelfth International Conference on Learning Representations*, 2023.

Margaret Mitchell, Avijit Ghosh, Alexandra Sasha Luccioni, and Giada Pistilli. Fully autonomous ai agents should not be developed. *arXiv preprint arXiv:2502.02649*, 2025.

Raja Parasuraman. Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43(7):931–951, 2000.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Series in Artificial Intelligence. Prentice-Hall, Englewood Cliffs, NJ, 1995.

Chirag Shah and Ryen W. White. Agents are not enough. 2024. URL https://arxiv.org/abs/2412.16241.

Yijia Shao, Yucheng Jiang, Theodore Kanell, Peter Xu, Omar Khattab, and Monica Lam. Assisting in writing Wikipedia-like articles from scratch with large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6252–6278, Mexico City, Mexico, June 2024a. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.347. URL https://aclanthology.org/2024.naacl-long.347/.

Yijia Shao, Vinay Samuel, Yucheng Jiang, John Yang, and Diyi Yang. Collaborative gym: A framework for enabling and evaluating human-agent collaboration. *arXiv preprint arXiv:2412.15701*, 2024b.

Theodore Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2023.

Maja Svanberg, Wensu Li, Martin Fleming, Brian Goehring, and Neil Thompson. Beyond ai exposure: Which tasks are cost-effective to automate with computer vision? *Available at SSRN 4700751*, 2024.

Suzanne Tsacoumis and Shannon Willison. O* net analyst occupational skill ratings: Procedures. *Alexandria, VA: Human Resources Research Organization*, 2010.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024a.

Xingyao Wang, Boxuan Li, Yufan Song, Frank F. Xu, Xiangru Tang, Mingchen Zhuge, Jiayi Pan, Yueqi Song, Bowen Li, Jaskirat Singh, Hoang H. Tran, Fuqiang Li, Ren Ma, Mingzhang Zheng, Bill Qian, Yanjun Shao, Niklas Muennighoff, Yizhe Zhang, Binyuan Hui, Junyang Lin, Robert Brennan, Hao Peng, Heng Ji, and Graham Neubig. OpenHands: An Open Platform for AI Software Developers as Generalist Agents, 2024b. URL https://arxiv.org/abs/2407.16741.

Michael Wooldridge and Nicholas R Jennings. Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152, 1995.

Tianbao Xie, Danyang Zhang, Jixuan Chen, Xiaochuan Li, Siheng Zhao, Ruisheng Cao, Toh J Hua, Zhoujun Cheng, Dongchan Shin, Fangyu Lei, et al. Osworld: Benchmarking multimodal agents for open-ended tasks in real computer environments. *Advances in Neural Information Processing Systems*, 37:52040–52094, 2024.

John Yang, Carlos E Jimenez, Alexander Wettig, Kilian Lieret, Shunyu Yao, Karthik R Narasimhan, and Ofir Press. Swe-agent: Agent-computer interfaces enable automated software engineering. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Shunyu Yao, Noah Shinn, Pedram Razavi, and Karthik Narasimhan. τ-bench: A benchmark for tool-agent-user interaction in real-world domains, 2024. URL https://arxiv.org/abs/2406.12045.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. Wildchat: 1m chatgpt interaction logs in the wild. *arXiv preprint arXiv:2405.01470*, 2024.

Appendix

Table of Contents

A Survey Details

We instantiate our auditing framework (see §2) with an audio-enhanced, semi-structured survey interface to collect first-hand data from domain workers who actually perform those tasks. Our survey is structured as follows:

- A **Mini-interview Section** designed to explore participants' work process and perspectives on the role of AI agents in their work. This section consists of five open-ended questions, allowing participants to share their thoughts freely and edit their audio transcripts in real time:

  A1 Could you please briefly describe what you do for your work?

  A2 What tasks do you typically do for your work? Think about this question based on the time you spend on each of them.

  A3 Please tell us more about the tools or software you use for these tasks. Please try to sort them by usage frequency.

  A4 For the three tasks that you spend the most of your time on, could you walk us through your process of completing each of them?

  A5 How do you envision using AI in your daily work?

- A **Task Rating Section** assessing both automation desire ($A_w(t)$) and the desired Human Agency Scale level ($H_w(t)$) for tasks associated with the participant's occupation. For each task, participants respond to a series of structured questions. Items T.I3, T.A1–A3, and T.C1–C5 are all rated on a 5-point Likert scale:

  TASK FAMILARITY QUESTIONS:

  T.I1 Have you done this task before?
  (Yes/No; if "No", remaining questions are skipped.)

  T.I2a With respect to this task, I consider myself a...
  (Novice, Average, Expert)

  T.I2b How much time do you spend on this task in your daily work schedule?
  (10%, 30%, 50%, 70%, 100%)

  T.I3 How closely is this task related to your core skills or unique strengths that are essential to your job?

  AUTOMATION DESIRE RATING QUESTIONS:

  T.A1 If an AI can do this task for you completely, how worried would you be that your job will be replaced?

  T.A2 Without thinking about salary, how much do you enjoy doing this task?

  T.A3 If an AI can do this task for you completely, how much do you want an AI to do it for you?

  T.A4 Why would you like this task to be automated by AI? (Shown if T.A3 $\geq$ 3; multi-select options):
  - Automating this task would free up my time for higher-value work.
  - This task is repetitive or tedious.
  - Automating this task would improve the quality of my work.
  - The task is stressful or mentally draining.
  - This task is complicated or difficult.
  - Automating this task would help me scale and handle higher output.

  HUMAN AGENCY SCALE RATING QUESTIONS:

  T.C1 How much does this task require taking physical actions or physical labor?

T.C2 How much does this task require dealing with uncertainty or making high-stake decisions?

T.C3 How much does this task require specific domain expertise (such as specialized knowledge, unspoken wisdom, or insights gained through experience)?

T.C4 How much does this task depend on interpersonal communication or empathy?

T.C5 If AI were to assist in this task, how much of your collaboration would be needed to complete this task effectively? (References to H1 to H5 are provided)

T.C6 Why would collaboration be needed for this task? (Shown if T.C5 $\geq$ 3; multi-select options):
   * This task requires physical actions.
   * This task involves making high-stake decisions which I would like to control.
   * This task requires specific domain knowledge.
   * The task involves nuanced communication or interpersonal skills.
   * The task needs validation or oversight to ensure quality.
   * The task is dynamic and requires adapting to changing circumstances.
   * The task has ethical, sensitive, or subjective aspects.

- A **Demographic Question Section** where we ask about information on age, gender, race, income, education, years of experience in occupation, attitude towards AI, zip code, political orientation, and details about their familiarity with LLMs and how they currently use them (*i.e.*, types of usage and frequency).

B Robustness Analysis

B.1 Annotation Agreement of AI Expert Assessments

AI experts with practical R&D experience provided ratings for current automation capability $A_e(t)$ and feasible human-agency levels $H_e(t)$. To ensure high-quality annotations, we applied the following controls:

- **Expert qualifications.** Each expert satisfied at least one of:
    1. Current PhD student specializing in NLP, large language models, or AI agents.
    2. PhD in Computer Science with demonstrated expertise in AI, LLMs, or agentic systems.
    3. Industry practitioner (*e.g.*, machine learning engineer or research scientist) with hands-on experience in LLMs and agentic systems.

    In total, 52 experts were recruited from institutions including Stanford University, MIT, Google, and xAI, *etc.*.

- **Assessment protocol.** Every task t was independently assessed by at least two experts, with additional reviews to ensure that $A_e(t)$ and $H_e(t)$ ratings exhibit a standard deviation ≤ 1.

Inter-annotator agreement, measured by Krippendorff's α, was 0.539 for $A_e(t)$ and 0.511 for $H_e(t)$.

B.2 Mixed-Effects Model Regression on Worker Responses

To assess the extent to which workers' automation desire ratings reflect intrinsic task properties rather than individual demographics, we fitted a linear mixed-effects model of the form

$$y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{k,ij} + u_j + \varepsilon_{ij},$$

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2),$$

Variable	Coef.	Std.Err.	z	P>|z|	Variable	Coef.	Std.Err.	z	P>|z|
Intercept	2.736	0.224	12.208	0.000	llm_usage_by_type__edit[T.Monthly]	0.023	0.065	0.352	0.725
gender[T.Male]	0.070	0.037	1.909	0.056	llm_usage_by_type__edit[T.Never]	-0.185	0.076	-2.447	0.014
gender[T.Other]	-0.210	0.356	-0.591	0.555	llm_usage_by_type__edit[T.Weekly]	-0.094	0.047	-2.009	0.045
gender[T.Prefer not to say]	0.153	0.141	1.084	0.278	llm_usage_by_type__idea_generation[T.Monthly]	-0.096	0.062	-1.545	0.122
education[T.Bachelor's Degree]	0.022	0.081	0.268	0.789	llm_usage_by_type__idea_generation[T.Never]	0.072	0.076	0.952	0.341
education[T.Doctorate (e.g., PhD)]	0.236	0.113	2.086	0.037	llm_usage_by_type__idea_generation[T.Weekly]	-0.014	0.049	-0.279	0.781
education[T.High School]	0.037	0.117	0.319	0.750	llm_usage_by_type__communication[T.Monthly]	-0.042	0.066	-0.641	0.521
education[T.Master's Degree]	0.121	0.084	1.436	0.151	llm_usage_by_type__communication[T.Never]	-0.052	0.067	-0.785	0.432
education[T.Prefer not to say]	0.157	0.221	0.711	0.477	llm_usage_by_type__communication[T.Weekly]	-0.051	0.047	-1.084	0.278
education[T.Professional Degree (e.g., MD, JD)]	-0.087	0.130	-0.670	0.503	llm_usage_by_type__analysis[T.Monthly]	0.078	0.067	1.158	0.247
education[T.Some College, No Degree]	-0.125	0.093	-1.340	0.180	llm_usage_by_type__analysis[T.Never]	-0.013	0.080	-0.165	0.869
experience[T.3-5 years]	0.100	0.060	1.662	0.097	llm_usage_by_type__analysis[T.Weekly]	0.126	0.052	2.415	0.016
experience[T.6-10 years]	0.141	0.065	2.180	0.029	llm_usage_by_type__decision[T.Monthly]	0.022	0.066	0.329	0.742
experience[T.Less than 1 year]	-0.073	0.120	-0.610	0.542	llm_usage_by_type__decision[T.Never]	-0.085	0.077	-1.107	0.268
experience[T.More than 10 years]	0.229	0.067	3.415	0.001	llm_usage_by_type__decision[T.Weekly]	-0.122	0.054	-2.268	0.023
llm_familiarity[T.I have some experience using them.]	-0.266	0.154	-1.726	0.084	llm_usage_by_type__coding[T.Monthly]	-0.021	0.064	-0.328	0.743
llm_familiarity[T.I use them regularly.]	-0.266	0.156	-1.711	0.087	llm_usage_by_type__coding[T.Never]	0.057	0.067	0.850	0.396
llm_familiarity[T.No, I've never heard of them.]	1.370	0.611	2.241	0.025	llm_usage_by_type__coding[T.Weekly]	-0.130	0.059	-2.228	0.026
llm_use_in_work[T.Yes, I use them every day in my work.]	0.077	0.053	1.457	0.145	llm_usage_by_type__system_design[T.Monthly]	-0.171	0.073	-2.334	0.020
llm_use_in_work[T.Yes, I use them every week in my work.]	0.025	0.051	0.493	0.622	llm_usage_by_type__system_design[T.Never]	-0.173	0.073	-2.362	0.018
race[T.Black]	-0.077	0.069	-1.127	0.260	llm_usage_by_type__system_design[T.Weekly]	-0.131	0.070	-1.858	0.063
race[T.Hispanic]	-0.024	0.093	-0.260	0.795	llm_usage_by_type__data_processing[T.Monthly]	-0.026	0.060	-0.430	0.667
race[T.Native American]	-0.209	0.166	-1.257	0.209	llm_usage_by_type__data_processing[T.Never]	0.099	0.061	1.633	0.102
race[T.Other]	0.233	0.088	2.657	0.008	llm_usage_by_type__data_processing[T.Weekly]	0.026	0.052	0.494	0.622
race[T.White]	-0.139	0.062	-2.226	0.026	ai_tedious_work_attitude[T.Somewhat agree]	0.385	0.096	3.991	0.000
income[T.165K-209K]	0.209	0.080	2.621	0.009	ai_tedious_work_attitude[T.Somewhat disagree]	0.538	0.128	4.206	0.000
income[T.209K-529K]	0.212	0.102	2.076	0.038	ai_tedious_work_attitude[T.Strongly agree]	0.685	0.097	7.098	0.000
income[T.30-60K]	0.275	0.063	4.332	0.000	ai_tedious_work_attitude[T.Strongly disagree]	0.439	0.126	3.482	0.000
income[T.529K+]	0.687	0.181	3.792	0.000	ai_job_importance_attitude[T.Somewhat agree]	-0.027	0.050	-0.534	0.593
income[T.60-86K]	0.113	0.066	1.711	0.087	ai_job_importance_attitude[T.Somewhat disagree]	-0.040	0.056	-0.720	0.472
income[T.86K-165K]	0.089	0.064	1.403	0.161	ai_job_importance_attitude[T.Strongly agree]	-0.087	0.065	-1.351	0.177
income[T.Prefer not to say]	-0.035	0.107	-0.329	0.742	ai_job_importance_attitude[T.Strongly disagree]	-0.012	0.071	-0.172	0.864
political_affiliation[T.Green Party]	-0.198	1.157	-0.171	0.864	ai_daily_interest_attitude[T.Somewhat agree]	0.128	0.074	1.724	0.085
political_affiliation[T.Independent]	-0.074	0.053	-1.403	0.161	ai_daily_interest_attitude[T.Somewhat disagree]	0.027	0.103	0.262	0.793
political_affiliation[T.Libertarian]	0.193	0.125	1.539	0.124	ai_daily_interest_attitude[T.Strongly agree]	0.415	0.078	5.308	0.000
political_affiliation[T.No political affliation]	0.063	0.057	1.092	0.275	ai_daily_interest_attitude[T.Strongly disagree]	0.069	0.109	0.633	0.526
political_affiliation[T.Other]	0.208	0.157	1.328	0.184	ai_suffering_attitude[T.Somewhat agree]	-0.299	0.058	-5.191	0.000
political_affiliation[T.Prefer not to answer]	0.136	0.070	1.942	0.052	ai_suffering_attitude[T.Somewhat disagree]	-0.047	0.055	-0.852	0.394
political_affiliation[T.Republican]	0.054	0.048	1.140	0.254	ai_suffering_attitude[T.Strongly agree]	-0.438	0.074	-5.928	0.000
llm_usage_by_type__information_access[T.Monthly]	-0.082	0.065	-1.256	0.209	ai_suffering_attitude[T.Strongly disagree]	-0.023	0.059	-0.391	0.696
llm_usage_by_type__information_access[T.Never]	-0.125	0.077	-1.624	0.104	age	-0.003	0.002	-2.033	0.042
llm_usage_by_type__information_access[T.Weekly]	0.098	0.046	2.149	0.032					

Table 1: Fixed-effects estimates from the mixed-effects regression predicting automation desire ratings.

where y_{ij} is the automation desire rating provided by worker i on task j, $X_{k,ij}$ are the K demographic and attitude covariates (age, gender, education, experience, LLM familiarity/use, income, political affiliation, LLM-usage subtypes, and AI-attitude scales), β_k their fixed effects, u_j a task-specific random intercept, and ε_{ij} the residual error. The model was estimated by REML using the `statsmodels` MixedLM interface in Python.

The fitted model (see Table 1) yielded the following variance-component estimates:

$$\widehat{\sigma}_u^2 = 0.066, \qquad \widehat{\sigma}_\varepsilon^2 = 1.254,$$

implying an intraclass correlation

$$\text{ICC} = \frac{\widehat{\sigma}_u^2}{\widehat{\sigma}_u^2 + \widehat{\sigma}_\varepsilon^2} = 0.050,$$

i.e., roughly 5% of total variance in automation desire is attributable to between-task differences. This "small-to-moderate" ICC confirms that task-level properties carry a real signal in workers' automation desire ratings.

Among fixed effects, higher educational attainment ("Doctorate" vs. "Bachelor's": $\hat{\beta} = 0.236$, $p = 0.037$) and greater work experience (">10 years" vs. "1–2 years": $\hat{\beta} = 0.229$, $p < 0.01$) were associated with increased automation desire. Attitudinal scales also showed significant associations: Strong agreement that "AI relieves tedious work" predicted higher desire ($\hat{\beta} = 0.685$, $p < 0.001$), while

stronger "AI suffering" attitudes predicted lower desire ($\hat{\beta} = -0.438, p < 0.001$). Income levels were positively related to automation desire (*e.g.* "$529K+" vs. "$0–30K": $\hat{\beta} = 0.687, p < 0.001$).

Together, these results indicate that, after controlling for a broad array of individual differences, task identity still explains a meaningful fraction of variance in automation desire. In §3, we use the average ratings for analysis.

C   Usage of External Data and Resources

C.1   Occupational Information Network (O*NET)

We source occupational tasks in this study from O*NET (version 29.2) Task Statements[3]. The O*NET database is a regularly updated database containing information about occupations across the United States. O*NET maps occupations to knowledge, skills, and abilities on different levels of granularity, as well as to tasks and detailed work activities. In O*NET, tasks are specific work activities that can be unique for each occupation. In total, there are 18,796 task statements spanning across 923 occupations.

Each task statement is associated with O*NET-SOC Code, Title (*i.e.*, occupation), and Task Type (*i.e.*, "Core"/"Supplementary"). Moreover, O*NET provides annotations of task categories based on the frequency of the task in seven categories ("Yearly or less", " More than yearly", "More than monthly", "More than weekly", "Daily", " Several times daily", " Hourly or more")[4]. As this work focuses on digital AI agents, we filter these task statements based on the following criteria:

1. The occupation mainly involves using computers in its work as judged by `gpt-4o`.

2. The task can be finished on the computer as judged by `gpt-4o`

3. The task does not miss annotation for "Core"/"Supplementary".

4. The task will be done more than monthly.

After filtering, there are 2,131 tasks remaining, spanning across 287 occupations.

> **Prompt for Filtering Occupations**
>
> Does this job mainly involve using computers?
> Answer format: "Yes" or "No"
> Occupation: {occupation}

> **Prompt for Filtering Occupational Tasks**
>
> For {occupation}, is it possible to finish its work-related task on a computer?
> Answer format: "Yes" or "No"
> Task: {task}

C.2   Occupational Employment and Wage Statistics by U.S. Bureau of Labor Statistics

We use occupational employment and wage statistics from the U.S. Bureau of Labor Statistics (BLS) to contextualize our findings with economic data. Specifically, we draw on data from the BLS May 2024

---

[3]https://www.onetcenter.org/dictionary/29.2/excel/task_statements.html
[4]https://www.onetcenter.org/dictionary/29.2/excel/task_categories.html

Occupational Employment and Wage Statistics Query System[5] to obtain the "Annual Mean Wage" and "Employment" (*i.e.*, number of employees) fields for each occupation in our database. These fields, combined with the collected first-hand data in WORKBank, inform the analysis presented in Figure 7.

C.3 Claude.ai Usage Data

To shed light on the relationship between current large language model (LLM) usage and future worker desires, we compare WORKBank automation desires with existing Claude.ai usage data from the Anthropic Economic Index Handa et al. (2025). The Anthropic dataset reports Claude usage at the task level, following O*NET task definitions. These standardized definitions allow us to directly map tasks from the Anthropic dataset to corresponding tasks in the WORKBank database, enabling a structured comparison across both sources. Our data shows that the top 10 occupations with the highest average automation desire represent only 1.26% of total usage (Figure 4 **c**).

C.4 Y Combinator (YC) Company and AI Agent Research Paper Data

We collect data on Y Combinator (YC) companies and AI agent research papers to assess how current investment and research efforts align with the desire–capability landscape revealed by the WORKBank database (Figure 5). To enable this analysis, we developed an LLM-assisted pipeline that systematically identifies and maps relevant YC startups and academic publications to specific occupational tasks in the WORKBank database.

YC Company Collection Process　The full list of YC companies was retrieved on April 28, 2025, from the official YC website[6]. The initial dataset comprised 5,156 companies. Company descriptions were collected from their respective YC detail pages and filtered using an LLM-based process (`gpt-4.1-mini`). Each company description was assessed using a binary classification prompt (see Prompt for YC Company Classifier) to determine whether the company is AI-relevant. This process identified 1,723 AI-related companies.

AI Agent Research Paper Collection Process　Academic papers were obtained from the arXiv official website[7], with a submission cut-off date of April 24, 2025. Papers were first screened by keyword: their title or abstract must contain "language model" (case-insensitive) and either "agent" or "system." This yielded an initial set of 17,064 papers. This set was refined with `gpt-4.1-mini` using a checklist (see Prompt for AI Agent Paper Classifier) to verify that each paper: (i) describes tasks extending *beyond single-turn raw text completion*, (ii) presents an implemented pipeline, (iii) conducts task-level evaluations, and (iv) involves a realistic task scenario. This process produced a final selection of 1,222 papers. For each paper passing this filter, we performed an additional round of task extraction (see Prompt for Paper Task Extractor); the extracted task statement serves as the paper's representative description in the subsequent mapping step.

Task Mapping Process　For each YC company and AI agent paper identified through the above processes, we again employed `gpt-4.1-mini` to perform binary classification of their applicability to each occupational task in the WORKBank database. For YC company-to-task mapping, see Prompt for Company-to-Task Classifier; for agent paper-to-task mapping, see Prompt for Paper-to-Task Classifier.

---

[5] https://www.bls.gov/oes/tables.htm
[6] https://www.ycombinator.com/companies
[7] http://arxiv.org

Prompt for YC Company Classifier

You will be presented with a company description. Your job is to classify if the company is an AI related company or not.
An AI-related company is defined as a company that is involved in the research, development, or application of AI. Output a boolean value.
———

The description of the company: {company_description}
The boolean value indicating if the company is an AI-related company:

Prompt for AI Agent Paper Classifier

You will see a paper TITLE and ABSTRACT. Return True if the paper's main contribution is an LLM-driven AGENT SYSTEM, else False.

Agent-system criteria (must ALL hold):
1. Beyond single turn raw text completion - The LLM's output decides what action or module happens next (planner/controller role), beyond single turn raw text completion.
2. Implemented pipeline - A complete system is actually built and run, not merely proposed.
3. Task-level evaluation - The paper reports results on the entire system performing its task (automatic metrics or user studies).
4. Realistic task - The task matches a plausible real-world workflow or a credible simulated environment.

If ANY criterion is missing, output "False".
———

The title of the paper: {paper_title}
The abstract of the paper: {paper_abstract}
Whether the paper is an LLM-driven agent system:

Prompt for Paper Task Extractor

You will be given the TITLE and ABSTRACT of a research paper describing an agent system. Extract the core task the paper addresses and express it in one concise sentence that begins exactly with: "The paper proposes an agent system to solve the task of...". Your output should be only that sentence, capturing the primary objective of the system.
———

The title of the paper: {paper_title}
The abstract of the paper: {paper_abstract}
The core task the paper addresses, expressed in one concise sentence:

Prompt for Company-to-Task Classifier

You will receive a brief description of a company and its product/service and an indexed list of workflows. For each workflow, decide whether workers involved in that workflow are a primary or explicitly intended user group of the company's offering. If the link is merely incidental, indirect, or speculative, mark it False. When in doubt, default to False.

———

The description of the company: {company_description}

The list of workflows and their descriptions: {workflows}

The dictionary of occupations and whether they are the target users of the company:

Prompt for Paper-to-Task Classifier

You will receive a brief description of a task proposed in the paper and an indexed list of workflows. For each workflow, decide whether the task is related to the workflow and the research is applicable to the workflow. If the link is merely incidental, indirect, or speculative, mark it False. When in doubt, default to False.

———

The description of the task proposed in the paper: {task_description}

The list of workflows and their descriptions: {workflows}

The dictionary of workflows and whether they are related to the task:

D WORKBank Statistics

As detailed in Section 2.5, we retained only occupations with at least ten worker responses from January to May 2025, yielding 1,500 individual assessments across 104 occupations. In this section, we present detailed statistics on occupational coverage and participant demographics in the WORKBank database.

D.1 Full List of Included Occupations

Occupation (O*NET-SOC Title)	N	Occupation (O*NET-SOC Title)	N
Customer Service Representatives	53	Biostatisticians	12
Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products	35	Quality Control Analysts	12
Accountants and Auditors	31	Advertising and Promotions Managers	12
Clinical Research Coordinators	30	Budget Analysts	11
Medical and Health Services Managers	30	Public Relations Specialists	11
Computer Programmers	28	Financial Managers	11
Web Developers	27	Logistics Analysts	11
Computer and Information Systems Managers	27	Medical Transcriptionists	11
Computer Systems Analysts	25	Radiologists	11
Purchasing Managers	24	Eligibility Interviewers, Government Programs	11
Business Teachers, Postsecondary	24	Court, Municipal, and License Clerks	11
Information Technology Project Managers	24	Securities, Commodities, and Financial Services Sales Agents	11
Financial Quantitative Analysts	23	Sustainability Specialists	11
Insurance Claims and Policy Processing Clerks	21	Web Administrators	11
Computer and Information Research Scientists	20	Geographers	11
Legal Secretaries and Administrative Assistants	19	Management Analysts	11
Secretaries and Administrative Assistants, Except Legal, Medical, and Executive	18	Bioinformatics Scientists	11
Human Resources Specialists	18	News Analysts, Reporters, and Journalists	11
Human Resources Managers	17	Compliance Officers	11
Business Intelligence Analysts	17	Video Game Designers	11
Purchasing Agents, Except Wholesale, Retail, and Farm Products	17	Lawyers	11
Statisticians	16	Proofreaders and Copy Markers	10
Personal Financial Advisors	16	Architects, Except Landscape and Naval	10
Production, Planning, and Expediting Clerks	16	Art Directors	10
Computer User Support Specialists	16	Data Entry Keyers	10
Search Marketing Strategists	16	Public Safety Telecommunicators	10
Architectural and Civil Drafters	15	Producers and Directors	10
Media Technical Directors/Managers	15	Precision Agriculture Technicians	10
Editors	15	Credit Counselors	10
Training and Development Specialists	15	Tax Preparers	10
Transportation Planners	15	Health Informatics Specialists	10
Appraisers and Assessors of Real Estate	14	Aerospace Engineers	10
Fundraisers	14	Medical Secretaries and Administrative Assistants	10
Credit Analysts	14	Technical Writers	10
Writers and Authors	14	Reservation and Transportation Ticket Agents and Travel Clerks	10
Graphic Designers	14	Petroleum Engineers	10
Civil Engineers	14	Molecular and Cellular Biologists	10
Information Security Analysts	14	Social Science Research Assistants	10
Online Merchants	14	Financial Examiners	10
Mathematicians	13	Telemarketers	10
Travel Agents	13	Payroll and Timekeeping Clerks	10
Photographers	13	Quality Control Systems Managers	10
Software Quality Assurance Analysts and Testers	13	Judicial Law Clerks	10
Bookkeeping, Accounting, and Auditing Clerks	13	Database Administrators	10
Desktop Publishers	12	Power Distributors and Dispatchers	10
Cost Estimators	12	Energy Engineers, Except Wind and Solar	10
Regulatory Affairs Managers	12	Network and Computer Systems Administrators	10
Mechanical Engineers	12	Librarians and Media Collections Specialists	10
Supply Chain Managers	12	Treasurers and Controllers	10
Clinical Data Managers	12	Biofuels Production Managers	10
Computer Systems Engineers/Architects	12	Mechanical Engineering Technologists and Technicians	10
Loan Officers	12	Computer Network Support Specialists	10

Table 2: List of occupations covered in the WORKBank database and the number of survey participants from each occupation.
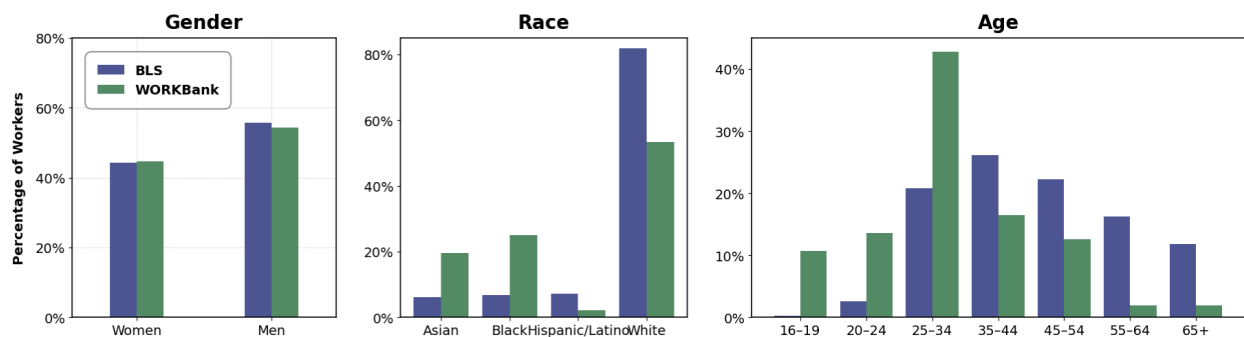
Figure 8: Demographic distribution of worker participants in our study compared to U.S. workforce demographics for the same set of occupations covered in the WORKBank database, based on data from the Bureau of Labor Statistics.

D.2   Coverage of the U.S. Workforce

We evaluate the representativeness of WORKBank by comparing its sector-level distribution with U.S. workforce data from the Bureau of Labor Statistics[8]. Figure 3 presents a breakdown of workforce coverage by sector, contrasting our database with the full U.S. workforce (**a**) and with the workforce restricted to the 104 occupations included in our database (**b**). Overall, the comparisons suggest that our database captures a broad and representative cross-section of the U.S. workforce at the sector level.

D.3   Domain Worker Demographic Information

We compare the demographic profile of domain workers in WORKBank with that of the U.S. workforce. U.S. workforce demographic data are sourced from the 2024 Annual Averages of the Bureau of Labor Statistics (BLS) Current Population Survey[9]. To ensure a fair comparison, we filter the BLS data to include only the 104 occupations represented in our database. Figure 8 shows the breakdown across key demographic dimensions. Our participant pool has a comprehensive demographic coverage, with a tendency toward younger age groups.

---

[8]https://www.bls.gov/oes/tables.htm
[9]https://www.bls.gov/cps/tables.htm

E Additional Results

E.1 Top 20 Tasks Workers Want Automated

Task	Average Automation Desire
Tax Preparers: Schedule appointments with clients.	5.00
Public Safety Telecommunicators: Maintain files of information relating to emergency calls, such as personnel rosters and emergency call-out and pager files.	4.67
Payroll and Timekeeping Clerks: Issue and record adjustments to pay related to previous errors or retroactive increases.	4.60
Desktop Publishers: Convert various types of files for printing or for the Internet, using computer software	4.50
Online Merchants: Create or maintain database of customer accounts.	4.50
Quality Control Systems Managers: Direct the tracking of defects, test results, or other regularly reported quality control data.	4.50
Statisticians: Report results of statistical analyses, including information in the form of graphs, charts, and tables.	4.50
Computer User Support Specialists: Maintain records of daily data communication transactions, problems and remedial actions taken, or installation activities.	4.50
Online Merchants: Calculate revenue, sales, and expenses, using financial accounting or spreadsheet software.	4.40
Data Entry Keyers: Store completed documents in appropriate locations.	4.33
Petroleum Engineers: Maintain records of drilling and production operations.	4.33
Logistics Analysts: Apply analytic methods or tools to understand, predict, or control logistics operations or processes.	4.33
Court, Municipal, and License Clerks: Instruct parties about timing of court appearances.	4.33
Data Entry Keyers: Maintain logs of activities and completed work.	4.25
Compliance Officers: Prepare correspondence to inform concerned parties of licensing decisions or appeals processes.	4.25
Web Developers: Back up files from Web sites to local directories for instant recovery in case of problems.	4.20
Web Administrators: Back up or modify applications and related data to provide for disaster recovery.	4.20
Bioinformatics Scientists: Manipulate publicly accessible, commercial, or proprietary genomic, proteomic, or post-genomic databases.	4.17
Network and Computer Systems Administrators: Perform routine network startup and shutdown procedures, and maintain control records.	4.17
Computer and Information Research Scientists: Approve, prepare, monitor, and adjust operational budgets.	4.17

Table 3: Top 20 tasks with highest average automation desire.

E.2 Bottom 20 Tasks Workers Want Automated

Task	Average Automation Desire
Reservation and Transportation Ticket Agents and Travel Clerks: Trace lost, delayed, or misdirected baggage for customers.	1.50
Logistics Analysts: Contact potential vendors to determine material availability.	1.50
Editors: Write text, such as stories, articles, editorials, or newsletters.	1.60
Reservation and Transportation Ticket Agents and Travel Clerks: Contact customers or travel agents to advise them of travel conveyance changes or to confirm reservations.	1.67
Video Game Designers: Provide feedback to designers and other colleagues regarding game design features.	1.67
Librarians and Media Collections Specialists: Code, classify, and catalog books, publications, films, audio-visual aids, and other library materials, based on subject matter or standard library classification systems.	1.67
Editors: Plan the contents of publications according to the publication's style, editorial policy, and publishing requirements.	1.67
Database Administrators: Write and code logical and physical database descriptions and specify identifiers of database to management system, or direct others in coding descriptions.	1.67
Graphic Designers: Key information into computer equipment to create layouts for client or supervisor.	1.67
Mechanical Engineering Technologists and Technicians: Calculate required capacities for equipment of proposed system to obtain specified performance and submit data to engineering personnel for approval.	1.67
Secretaries and Administrative Assistants, Except Legal, Medical, and Executive: Establish work procedures or schedules and keep track of the daily work of clerical staff.	1.67
Graphic Designers: Review final layouts and suggest improvements, as needed.	1.71
Graphic Designers: Prepare illustrations or rough sketches of material, discussing them with clients or supervisors and making necessary changes.	1.71
Mechanical Engineering Technologists and Technicians: Interpret engineering sketches, specifications, or drawings.	1.75
Accountants and Auditors: Prepare, examine, or analyze accounting records, financial statements, or other financial reports to assess accuracy, completeness, and conformance to reporting and procedural standards.	1.75
Editors: Allocate print space for story text, photos, and illustrations according to space parameters and copy significance, using knowledge of layout principles.	1.75
Producers and Directors: Cut and edit film or tape to integrate component parts into desired sequences.	1.75
Graphic Designers: Create designs, concepts, and sample layouts, based on knowledge of layout principles and esthetic design concepts.	1.78
Librarians and Media Collections Specialists: Locate unusual or unique information in response to specific requests.	1.80
Editors: Assign topics, events and stories to individual writers or reporters for coverage.	1.80

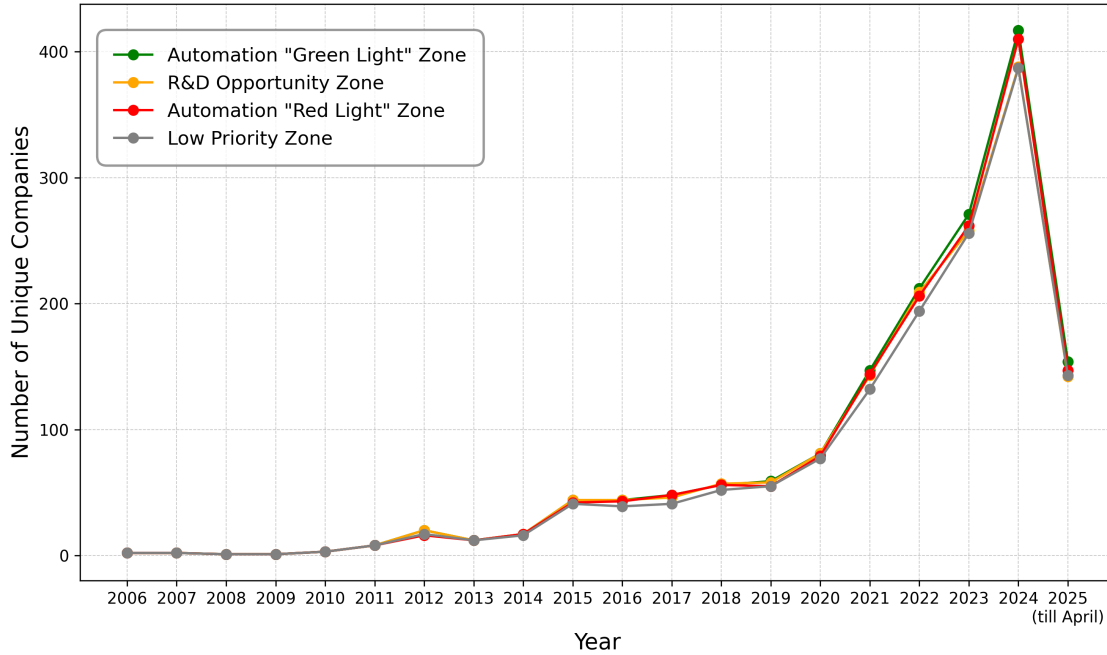Table 4: Bottom 20 tasks with lowest average automation desire.

Figure 9: **Number of unique newly funded Y Combinator companies mapped to each task zone in the automation desire–capability landscape (2006–2025, through April).** Despite the exponential growth in AI agent startups, the temporal trends across all four zones remain largely parallel. Notably, there is no disproportionate concentration in the Automation "Green Light" or R&D Opportunity Zones—areas that warrant greater attention.

E.3   Y Combinator Investment Patterns Across Task Zones

As discussed in §3.2, jointly considering the worker-rated automation desire $A_w(t)$ and expert-assessed technological capability $A_e(t)$ divide tasks in WORKBank into four zones: Automation "Green Light" Zone, Automation "Red Light" Zone, R&D Opportunity Zone, Low Priority Zone

Figure 9 illustrates the number of unique newly funded YC companies mapped to each task zone from 2006 to 2025 (till April). Despite the exponential growth in AI agent startups, the distribution across zones has remained relatively uniform over time. Notably, there is no disproportionate concentration in the Automation "Green Light" or R&D Opportunity Zones—areas that warrant greater attention. While the task zone classifications reflect a static snapshot and may not reflect the status in the past, the findings nonetheless suggest a misalignment between where investments are flowing and the joint perspective of both those developing the technology and the workers the technology shall aim to support.
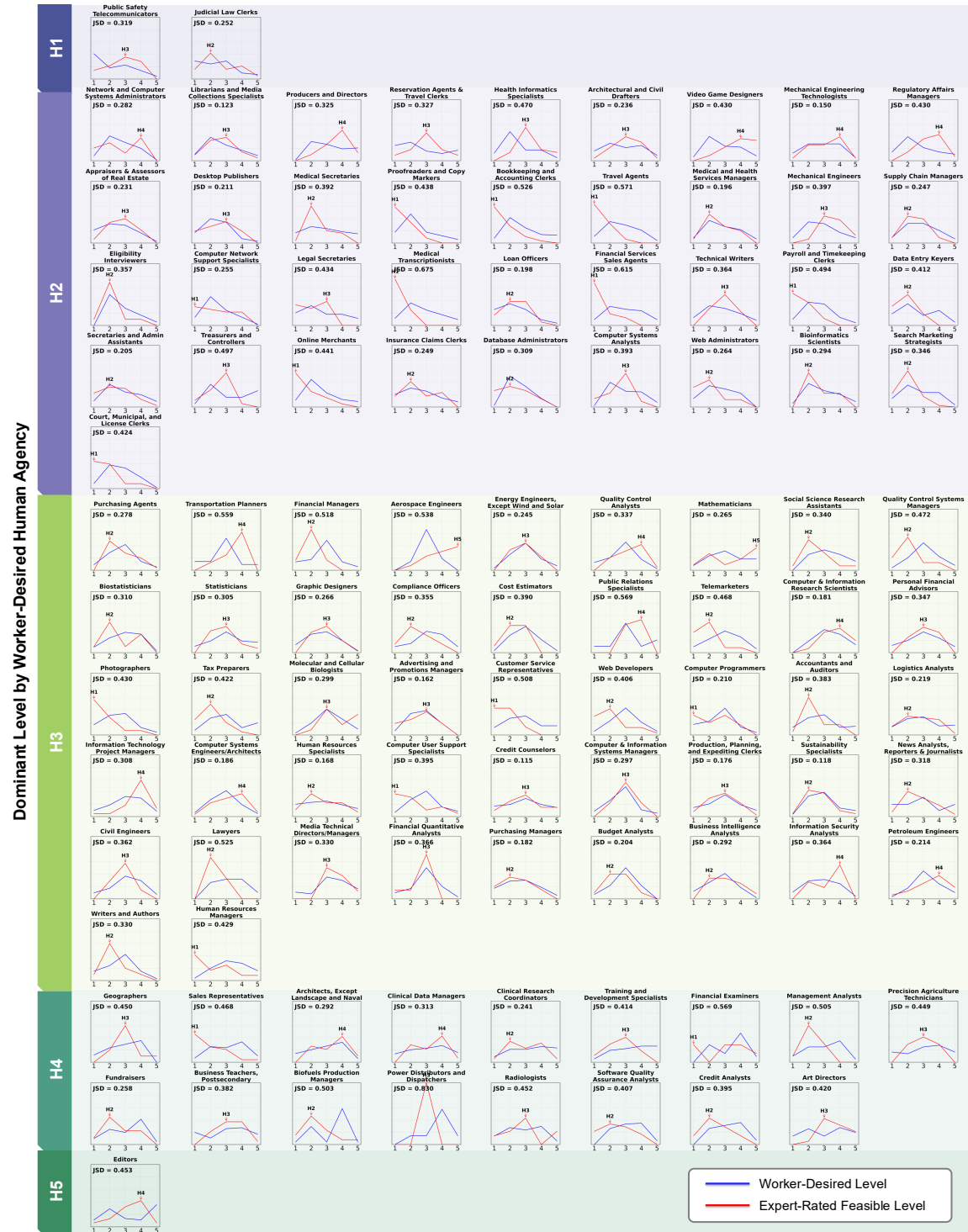
E.4 Full Human Agency Scale Results

Figure 10: **Distributions of Human Agency Scale (HAS) levels.** The Jensen–Shannon Divergence (JSD) quantifies the divergence between the distribution of worker-desired HAS levels ($H_w(t)$) and expert-assessed feasible HAS levels ($H_e(t)$).

E.5   Top 10 Occupations By Worker-Expert Discrepancies in HAS Ratings

Occupation	Dominant HAS Level By Worker Desire	Dominant HAS Level By AI Expert Assessment	JSD
Power Distributors and Dispatchers	H4	H3	0.830
Medical Transcriptionists	H2	H1	0.675
Securities, Commodities, and Financial Services Sales Agents	H2	H1	0.615
Travel Agents	H2	H1	0.571
Financial Examiners	H4	H1	0.569
Public Relations Specialists	H3	H4	0.569
Transportation Planners	H3	H4	0.559
Aerospace Engineers	H3	H5	0.538
Bookkeeping, Accounting, and Auditing Clerks	H2	H1	0.526
Lawyers	H3	H2	0.525

Table 5: Top 10 occupations with the largest discrepancies between the worker-desired Human Agency Scale levels $H_w(t)$ and AI expert-assessed feasible levels $H_e(t)$. The discrepancy is measured by the Jensen–Shannon divergence (JSD) between these two distributions, where distributions are computed based on all annotated tasks within each occupation.

E.6   Task to Skill Mapping

We use the O*NET database to map tasks to their related skills (Generalized Work Activities). In order to create broad groups of common skills across multiple tasks, O*NET defines three levels of work activities: Generalized (GWA), Intermediate (IWA), and Detailed Work Activities (DWA). Each task in the O*NET database is mapped to one or more DWA's, which then corresponds to exactly one GWA.

For example, the task "Compile financial data to prepare quarterly budget reports" is mapped to the DWA "Prepare financial documents, reports, or budgets." This DWA is then associated with the skill (GWA): "Documenting/Recording Information". Using this mapping, we are able to directly match tasks to their corresponding skills.

Here, we provide the full descriptions of the skills shown in Figure 7, ordered by decreasing required human agency (Tsacoumis and Willison, 2010). We exclude physical skills (*i.e.,* "Identifying Objects, Actions, and Events," "Performing General Physical Activities," "Controlling Machines and Processes," "Inspecting Equipment, Structures, or Materials," "Handling and Moving Objects," "Repairing and Maintaining Mechanical Equipment") and "Working with Computers," as the audit focuses on tasks that are likely to be exposed to digital AI agents.

1. **Organizing, Planning, and Prioritizing Work**: Developing specific goals and plans to prioritize, organize, and accomplish your work.

2. **Training and Teaching Others**: Identifying the educational needs of others, developing formal educational or training programs or classes, and teaching or instructing others.

3. **Staffing Organizational Units**: Recruiting, interviewing, selecting, hiring, and promoting employees in an organization.

4. **Updating and Using Relevant Knowledge**: Keeping up-to-date technically and applying new knowledge to your job.

5. **Developing Objectives and Strategies**: Establishing long-range objectives and specifying the strategies and actions to achieve them.

6. **Guiding, Directing, and Motivating Subordinates**: Providing guidance and direction to subordinates, including setting performance standards and monitoring performance.

7. **Judging the Qualities of Objects, Services, or People**: Assessing the value, importance, or quality of things or people.

8. **Communicating with Supervisors, Peers, or Subordinates**: Providing information to supervisors, coworkers, and subordinates by telephone, in written form, e-mail, or in person.

9. **Providing Consultation and Advice to Others**: Providing guidance and expert advice to management or other groups on technical, systems-, or process-related topics.

10. **Thinking Creatively**: Developing, designing, or creating new applications, ideas, relationships, systems, or products, including artistic contributions.

11. **Interpreting the Meaning of Information for Others**: Translating or explaining what information means and how it can be used.

12. **Making Decisions and Solving Problems**: Analyzing information and evaluating results to choose the best solution and solve problems.

13. **Monitoring Processes, Materials, or Surroundings**: Monitoring and reviewing information from materials, events, or the environment, to detect or assess problems.

14. **Assisting and Caring for Others**: Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.

15. **Getting Information**: Observing, receiving, and otherwise obtaining information from all relevant sources.

16. **Monitoring and Controlling Resources**: Monitoring and controlling resources and overseeing the spending of money.

17. **Analyzing Data or Information**: Identifying the underlying principles, reasons, or facts of information by breaking down information or data into separate parts.

18. **Selling or Influencing Others**: Convincing others to buy merchandise/goods or to otherwise change their minds or actions.

19. **Documenting/Recording Information**: Entering, transcribing, recording, storing, of maintaining information in written or electronic/magnetic form.

Concept	Percentage of Summaries
Lack of trust in accuracy, reliability or capability	45.0%
Fear of job replacement	23.0%
Lack of AI's human qualities or capabilities	16.3%
AI not applicable or useful to specific work	15.6%

Table 6: Identified concepts with the seed prompt "The top most common fears that workers have about AI automation in their work".

20. **Evaluating Information to Determine Compliance with Standards**: Using relevant information and individual judgment to determine whether events or processes comply with laws, regulations, or standards.

21. **Communicating with People Outside the Organization**: Communicating with people outside the organization, representing the organization to customers, the public, government, and other external sources. The information can be exchanged in person, in writing, or by telephone or e-mail.

22. **Processing Information**: Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.

23. **Estimating the Quantifiable Characteristics of Products, Events, or Information**: Estimating sizes, distances, and quantities; or determining time, costs, resources, or materials needed to perform a work activity.

24. **Performing Administrative Activities**: Performing day-to-day administrative tasks such as maintaining information files and processing paperwork.

25. **Performing for or Working Directly with the Public**: Performing for people or dealing directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests.

26. **Scheduling Work and Activities**: Scheduling events, programs, and activities, as well as the work of others.

27. **Establishing and Maintaining Interpersonal Relationships**: Developing constructive and cooperative working relationships with others, and maintaining them over time.

F   Audio Response Analysis

F.1   LLM-based Topic Modeling

To analyze our audio transcripts collected from workers, we use LLooM (Lam et al., 2024), an LLM-based topic modeling tool that takes unstructured text to extract high-level concepts.

Using LlooM, we first apply a distillation step with `Claude 3 Opus` on the audio transcripts. This step filters the text to retain only the most relevant quotes based on a seed

Concept	Percentage of Summaries
Role-based support	23.1%
Assistantship	23.0%
Automation	16.5%
Separation of Tasks between AI and Humans	12.0%

Table 7: Identified concepts with the seed prompt "If workers want AI to help, what type of imagined partnership with AI do they prefer".

prompt and performs a summarization of the filtered text. Subsequently, we cluster these summarized quotes using the same LLM to form groups of related concepts. We manually inspect the clusters to ensure the concepts are distinct and merge any overlapping concepts as necessary. Table 6 and Table 7 present the top concept groups identified for two seed prompts, corresponding to the analyses in §3.1 and §3.3, respectively.

F.2 Audio Response Examples

F.2.1 Full Transcripts for Direct Quotes in §3.1

Art Director with 6–10 Years of Experience:

```
 So [my work is] basically just first of all looking through all the tasks in the
sauna at the start of the day.  And then once shoots and, you know, filmings are
happening, looking through the footage and pictures, making sure they adhere to
brand standards and guidelines and a cohesive voice.

 [I spend most of my time] Looking through imagery and making sure that it is
consistent and always deliverable, and making selects and things of that nature,
just establishing a cohesive tone.  [I use] So definitely Bridge, you know, Photo
Mechanic, Capture One, Photoshop, Asana for the tasks like I mentioned earlier,
you know, Gmail, things like that.  [For the detailed procedure,] Yeah, looking at
the imagery as it's flowing through during the shoot or, you know, filming if it's
video, and then from there going through and selecting and culling and, you know,
again, only sharing the best imagery that's cohesive.

 [For AI use,] I don't really, unless it's, you know, in some sort of minor
way to help the calling process become easier.  I don't want it to be used for
content creation.  If anything, I want it to be used for seamlessly maximizing
workflow and, you know, making things less repetitive and tedious and arduous with
workflow.  No content creation.
```

Art Director with 3-5 Years of Experience:

I manage some anime art projects as part of a company's public relations and community strategy for youth engagement. So I work with artists directly, manage projects and merchandise and tabling at events and all that fun stuff.

I do a lot of internal meetings just to make sure everyone's on the same page. It takes up a lot of my time. I also have to scope out projects, find artist to work with especially those we found on social media, figure out how to get in touch and work with them, work with community groups as well, do this type of stuff. And then I also help, well not directly do, but help assist in getting merchandise produced, including preparing artwork for like production and stuff like that.

I spend a lot of time in Microsoft Word, Microsoft Teams, Outlook, which, you know, my company uses Microsoft Office for everything. But then I also use, to communicate with artists, I use the apps that they use. So sometimes that's LINE, sometimes that's Discord, sometimes that's Twitter. And then I personally have my own notion for project management as well.

I spend a lot of time sculpting out projects, so I generally start brainstorming or collecting all my research, gather all my information into Notion first, and then put it into a Word document. The Word document's a little more formal, but I also like make sure that's still approachable to artists, and then I, you know, export that as a PDF. For communicating with artists, that depends on what they use, but most of them use Discord, so it's just back and forth. They send us a sketch, I send that off to my people for, like, feedback very quickly, and then I, you know, get back to them. I sometimes do have my own autonomy to, like, do the final say on what does work and what doesn't work. And then for at least getting everyone on the same page, I spend a lot of time Microsoft Teams. I obviously have to gather some meeting notes, like, write down some job stuff I want to talk about beforehand, make sure there's no surprises to people, it's just communicating and providing regular updates.

I would never use AI to like replace artists. I would be more for personal [project management] use, if anything, it's to summarize my tasks, for example, or things like improving my writing, using Apple's writing tools, where I can just revise my writing to be a little more concise, but I would never, let it brainstorm on my behalf, just because I find AI to be very poorly performing on those type of tasks.

Graphic Design with More Than 10 Years of Experience:

I do basically architecture presentation, like graphic design work, work like
typically. It's like layout design, which is organizing content, image or text
for AI storyline, diagram or infographics analysis diagram, render enhancement
like poster processing, 3D render for a polished look. I also do topography and
color scheme using professional fonts and color that align with the project's
theme. I also work with board composition, arranging plan, section, elevation
and perspectives. I also work digital and print formatting, which is ensuring
high quality output for print or physical brands.

So, I basically do architectural presentations. Graphic design work typically
includes layout design, diagram, etc. Later enhancement, like post-printing
3D renders for a polished work. I also give topography color scheme both for
precision, visibility, and formality. I basically use drawing for autocad and
after that for render I use Photoshop and Illustrator and for 3D render I use
Lumion. So basically I spend a lot of time drafting in AutoCAD. The most common
tasks just likely include creating floor plans, creating sections, creating
elevation, annotation and dimensioning drawings, organizing and managing the
layers.

AI can be a game-changer in data architect workflow, helping to improve
efficiency, accuracy and even creativity. But I create my design by myself. For
research, I use AI.

F.2.2 Transcripts from Occupations Exhibiting High Human Agency (HAS) Levels

As discussed in §3.3, the Human Agency Scale (HAS) spectrum reveals that very few
occupations are characterized by a dominant HAS Level 5 (indicating essential human
involvement). "Editors" is the only occupation where workers predominantly desire H5.
According to AI expert assessments, only "Mathematicians" and "Aerospace Engineers"
fall into this category. Below are representative transcripts for these occupations.

Editor with 3-5 Years of Experience:

I proofread and copy-edit marketing materials, mostly in the travel and tourism
sector. I also do some copywriting and script writing for different ad clients
and some light design work.

I look through flyers, brochures, other marketing materials, and I do several passes for mistakes in grammar, in consistency, in flow and clarity. And I make changes on the document, usually a PDF, and send them back to the client. They fix them, they send me another version, and I do several more passes until we've spent enough time and got it perfect. I mainly use Adobe products, PDFs in Adobe Reader. I also use Microsoft Word and some Adobe Suite products, mostly Illustrator and InDesign.

 For copy editing, I read through whatever material the client has sent me. I do a pass for basic grammar. I do a pass for clarity and flow, often changing the copy significantly to make it sound better. For proofreading, I go through the materials. Same thing, but with less of a view toward changing the copy and more toward finding errors in grammar and consistency and even design. For copywriting, I make an outline of my ideas for the project and complete a rough draft. Then I spend some time away from it and revise until I have a polished draft for the client.

 I'm resistant to using AI in my daily workflow. If I'm forced to use it, I would use it for basic grammar editing, but I would check each suggestion against my own knowledge very carefully and give it full consideration before adopting it as a change.

Editor With More Than 10 Years of Experience:

 So I work in a media company, [masked], and as an editor I make sure that all JavaScript that I'm going to print are formatted correctly, the colors are accurate, and there are no typos.

 So a lot of what I do involves sitting at a computer using productivity tools like Adobe Creative Suite, Canva, QuarkXPress, and using the Google Enterprise. So for email, document sharing, I use Google Docs quite a bit for my editing purposes, but I'll also receive files in PDF format. So just working with all the different tools on my computer to get my tasks done every day. So a lot of what I do involves sitting at a computer using productivity tools like Adobe Creative Suite, Canva, QuarkXPress, and using the Google Enterprise. So for email, document sharing, I use Google Docs quite a bit for my editing purposes, but I'll also receive files in PDF format. So just working with all the different tools on my computer to get my tasks done every day.

So one of the most frequent pieces of software I use is Adobe Acrobat, and that is really great for editing PDFs. The next most frequent software I use would be Google Docs. Receiving files through Google Docs, that's a great way to be able to provide updates and edits to annotate the files. And then I would say other tools like Adobe InDesign, Adobe Photoshop, QuarkXPress, Microsoft Publisher. Those are occasionally used, but that's really about all four of the frequently used programs, I would say. So one of the most frequent pieces of software I use is Adobe Acrobat, and that is really great for editing PDFs. The next most frequent software I use would be Google Docs. Receiving files through Google Docs, that's a great way to be able to provide updates and edits to annotate the files. And then I would say other tools like Adobe InDesign, Adobe Photoshop, QuarkXPress, Microsoft Publisher. Those are occasionally used, but that's really about all four of the frequently used programs, I would say.

So, when I'm reviewing PDF documents, I will use the markup tool to add comments and highlight certain sections to make sure that the wording is accurate, or if there's questions regarding the resolution of a photo, I can send that back to mark that up and say, hey, this needs to be a higher resolution photo, it won't print out correctly. So it's just a lot of manual review of every single file before it goes to print to make sure that everything is properly formatted, the colors are accurate, and it will reproduce correctly, just to make sure everything looks good for the customer. So, when I'm reviewing PDF documents, I will use the markup tool to add comments and highlight certain sections to make sure that the wording is accurate, or if there's questions regarding the resolution of a photo, I can send that back to mark that up and say, hey, this needs to be a higher resolution photo, it won't print out correctly. So it's just a lot of manual review of every single file before it goes to print to make sure that everything is properly formatted, the colors are accurate, and it will reproduce correctly, just to make sure everything looks good for the customer.

So, I'm using AI right now when it comes to email, so with the Google suite, there are Google Gemini tools that help with formatting emails. I can take a very simple format for content for email and then using that to expand those topics and make it more of a wordy email. But I would like to be able to use AI more in my proofreading and editing than I am right now, so probably within the next couple months I should be able to do that.

Mathematician With 3-5 Years of Experience:

I do number theory and algebraic geometry, mostly around long-length programs or categorical long-length programs. [In my daily work, I] read papers and write papers.

Solving a math problem, I don't know [whether there is any specific tool to use], just read papers and have an intuition of what the procedure of philosophy should be and work on it.

To be honest, I think [AI is] useless at this moment.

Mathematician With 3-5 Years of Experience:

I am studying geometric representation theory and categorical Langlands program. My work involves coming up the problem to work on and learning math tools to help me think of solutions.

I need to spend a lot of time reading papers and learning math tools. I also need to attend the seminar to find collaborators. Then I work on my problem. [In terms of tools,] I mainly use latex. I spend most of my time studying math. Papers in my field can have hundreds of pages – it takes a long time to understand and try to apply the technique.

At present, I don't think AI has any use for mathematicians, at least for DeepSeek and ChatGPT. One core question I am interested in is whether AI can come up with new stuffs that haven't been proposed before rather than solving problems people craft.

Mathematician With 6-10 Years of Experience:

I used to study number theory, in particular, p-adic Hodge theory in arithmetic geometry. Now, I work on the formalization of p-adic Hodge theory in Lean and also auto-formalization and auto theorem proving.

During formalization, I elaborate, generalize, and fill gaps in mathematical proofs. I design general fomalization frameworks and spend lots of time in Lean coding. Lean coding involves searching theorems, formalizing statements and filling in formalization details in the proof. The last part is the longest part. For auto formalization and formal theorem proving, I spend most of the time coding to establish the LLM's training pipeline and preparing data for the training. I use the interactive theorem prover Lean. I also use LeanSearch and other tools related to Lean to accelerate. I use Python for LLM training and use DeepSeek for coding and debugging.

[Here is a concrete example of my workflow:] I formalized a famous number theory definition, called the period rings of Fontaine. I first wrote down a detailed version of the mathematical statements and proofs I need. Splitting the whole formalization project into several smaller goals. For each smaller goal, I generalize and design suitable definitions and lemmas for formalization. Then I begin actual formalization using Lean. I first write down the definitions and state the theorems in Lean without proof. After this, I fill in the proofs backwards, searching the library for existing theorems to use and design patterns to mimic. During formalization, I revise the natural language proof from time to time.

 I think a primary AI tool could help me in filling in searching for theorems and design patterns during formalization. A stronger AI tool would do auto-formalization of theorem statements and provide suggestions in designs. An even stronger AI tool would be able to elaborate and fill gaps in mathematical proofs and autoformalize the human proofs. Additionally, an AI tool strong in coding, debugging, and software engineering would help a lot in coding.

Aerospace Engineer With 3-5 Years of Experience:

 I am an aeronautic engineer. I work in the aircraft maintenance, repairs, design aircraft. I work with [masked].

 We design aircraft, develop, test, and maintain aircrafts, and the systems that operate within Earth's atmosphere, such as airplane, helicopters, drones, and missiles, though we're not into missiles, though. Our work focuses on making aircraft, machines, very safe and efficient, capable of flight. We use Computer-Aided Design as a tool for Autodex, AutoCAD, Cartier, and Solidwork. And we use Computer Fluid Dynamics. It's ANSYS Fluent, STAR-CNC-MM. What else? We use Finite Element Analysis. It's a tool we use for Nastran.

 [In my opinion,] AI is going to be very awesome and it's going to make it very easier for us because most of the time, the main problem we have is detecting where the problem is in the engine, you know, so you have to do a lot of manual jobs and all that. So, but if we have AI, you can possibly tell in the dash cam or whatever, you know, you can possibly tell.

Aerospace Engineer With 1-2 Years of Experience:

 I'm a current undergraduate senior and prospective master's student in aerospace engineering, working in guidance, navigation, and controls, so like more simulation, computer programming side of aerospace engineering.

Most of what I do for work has traditionally been programming simulations to evaluate vehicle performance for orbital rockets. And so most of my tasks will be either building out a part of the simulation, programming new features or new testing, or kind of similar types of modeling of different subsystems of the rocket.

In general the tools or software that I use would be Visual Studio Code for the actual programming. The companies I've worked at have used project management tools like JIRA and Confluence. I think also just a lot of internet documentation is useful. And yeah, I very occasionally would use an AI tool like ChatGPT. Generally, my process would be to understand the requirements, which would involve talking to my manager, then kind of going about kind of like pre-reading or other types of information gathering necessary for the task, actual programming, and then like unit testing and other ways of forms of validation for the programming that I completed.

Honestly, I don't use AI too much in my current workflow. I think that the only time that it could come up would be if I'm running into some type of error or bug in my program that I can't find, or kind of a quick piece of code that I could look up how to do, but it's easier to just ask a AI model to generate. But honestly, I use it very, very infrequently.