Are We Automating the Joy Out of Work? Designing AI to Augment Work, Not Meaning

ANONYMOUS AUTHOR(S)

Prior work has mapped which workplace tasks are exposed to AI, but less is known about whether workers perceive these tasks as meaningful or as busywork. We examined: (1) which dimensions of meaningful work do workers associate with tasks exposed to AI; and (2) how do the traits of existing AI systems compare to the traits workers want. We surveyed workers and developers on a representative sample of 171 tasks and use language models to scale ratings to 10,131 tasks across all U.S. computer-assisted tasks. Worryingly, we find that tasks that workers associate with a sense of agency or happiness may be disproportionately exposed to AI. We also document HCI design gaps: developers report emphasizing politeness, strictness, and imagination in system design; by contrast, workers prefer systems that are straightforward, tolerant, and practical. To address these gaps, we call for AI whose design explicitly centers meaningful work and worker needs, proposing a five-part research agenda.

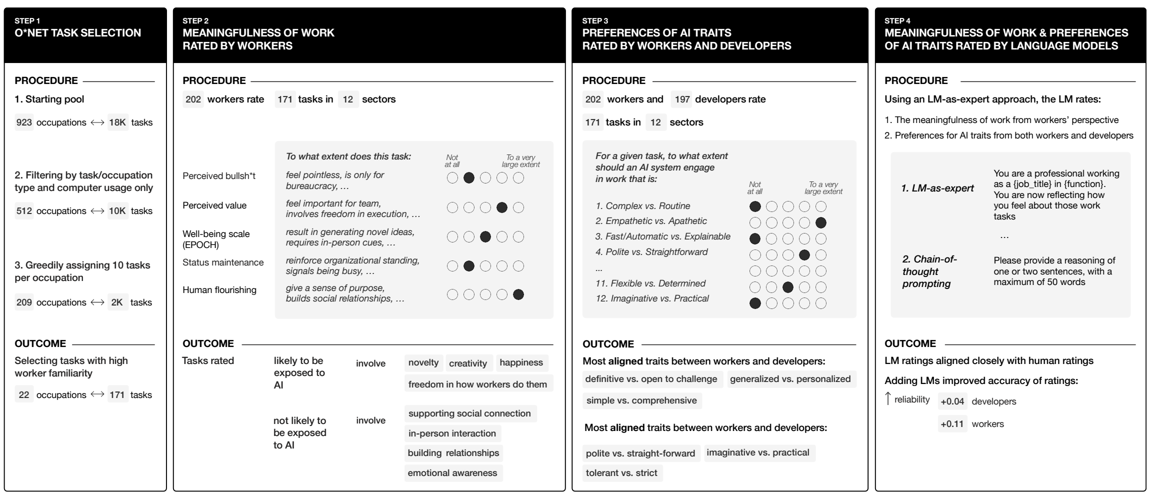


Fig. 1. **Overview of Study Design**: Worker and developer perspectives on meaningful work and AI system design in the U.S. labor force. *(Step 1)* Workplace tasks were restricted to those primarily completed on a computer and performed daily or weekly, then filtered by Prolific availability, AI Impact Index, and worker familiarity. *(Step 2)* Workers rated tasks across five dimensions: perceived bullsh*t, perceived value, well-being scale, status maintenance, and human flourishing. Tasks more likely exposed to AI scored higher on novelty, creativity, happiness, and freedom, while those less likely emphasized emotional awareness, in-person interaction, relationships, and social connection. *(Step 3)* Workers and developers rated which psychological traits an AI system should possess when augmenting tasks. When designing AI augmented tasks, developers emphasized polite, strict and imaginative systems whereas workers preferred straightforward, tolerant, and practical systems. *(Step 4)* LMs were prompted as experts to simulate worker and developer ratings, with moderate to high intra-class correlation with human responses.

Manuscript submitted to ACM

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**; • **Applied computing → Psychology**.

Additional Key Words and Phrases: Human-centered AI, Future of work, Automation and Augmentation, Meaningful Work, Value Alignment

ACM Reference Format:

Anonymous Author(s). 2026. Are We Automating the Joy Out of Work? Designing AI to Augment Work, Not Meaning. In *Proceedings of 2026 ACM CHI Conference on Human Factors in Computing Systems (CHI '26).* ACM, New York, NY, USA, 63 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Public debate about AI and occupations often focuses on job loss versus job growth. Some studies have predicted broad job displacement [32, 34, 91], while others have anticipated growth, with AI complementing workers in ways that are associated with higher productivity and the emergence of new roles [9, 17, 18, 23, 66, 67, 79]. These outcomes are not fixed, however, and are widely argued to vary with how teams design AI systems. In this paper, *AI* refers to software systems (e.g., LM tools, agents) that automate or augment computer-based tasks by generating, transforming, or routing information. By *teams*, we mean the broader set of actors involved in AI system development, including developers, UI/UX designers, product managers, and others who contribute to how these systems are built and used. Because the very possibility of drawing a clear distinction between automation and augmentation is debated in the literature [1, 7, 8, 31], with different works adopting different operationalizations (e.g., based on time saved by AI to do a task [28]), we clarify our terminology as follows: throughout, we use *AI exposure* [31], our main construct of interest, to refer to tasks that current or near-term AI systems could plausibly perform or substantially speed up, operationalized as those above the 75^{th} percentile of the distribution of the patent-based AI Impact Index [79]; *AI automation* for cases where AI can perform a task end-to-end with minimal or no human involvement; and *AI augmentation* for cases where AI supports or enhances human work, while humans retain primary responsibility and decision-making authority, with human involvement measured, for example, using the scale in [80].

To inform AI design, we examine: (1) Which tasks are likely to be exposed to AI? (2) How do workers evaluate these tasks in their daily work? (3) Do AI teams design systems that meet workers' needs? When these three aspects align, growth-oriented scenarios become more plausible; when any one fails, organizations may face greater waste and resistance, in the workplace and beyond.

The bulk of existing research lies in economics and has concentrated on the first aspect: identifying which tasks are exposed to AI, and measuring the resulting labor-market effects. This body of work forms the backbone of the literature: it maps technologies (e.g., via patents [9, 79]) to tasks, and quantifies how AI is associated with changes in occupations [43, 58, 79, 80]. A growing line of research has used LMs to annotate tasks [28]; for example, labeling a task as 'exposed' when AI is defined as enabling at least a 50% reduction in reported time to complete the task at equal or higher quality. By this definition, one study estimated that LMs are estimated to be relevant for the majority of tasks in just 1.8% of U.S. occupations, but this share is estimated to reach over 46% when AI is considered together with productivity software [28]. Yet, the same study also estimated that only about 1.86% of tasks are fully automatable without human oversight [28]. Therefore, while complete automation is uncommon, augmentation is expected to play the larger role [59]. In customer support, for example, the use of generative AI systems is estimated to be associated with roughly 14% higher measured agent productivity, on average, with increases of about 35% in the number of issues resolved per hour for less-experienced workers [80].

The second and third aspects: (2) how workers evaluate tasks exposed to AI, and (3) whether teams design AI to meet workers' needs, remain underexplored. These gaps present a promising research agenda for HCI. Recent work has begun to map where workers prefer human intervention, suggesting that workers report being comfortable with AI handling information-centric tasks, while preferring to focus on interpersonal and organizational work [80]. Yet, despite rich literature on meaningful work in management science [11, 39, 52, 56, 76, 77, 85] and on human-centered systems in HCI [3, 64, 65, 81], we still know little about which dimensions of meaningful work workers associate with tasks potentially exposed to AI, and whether AI teams design AI tools with the traits workers want. To address this gap, we asked two research questions:

(RQ1) Which dimensions of meaningful work do workers associate with tasks exposed to AI in their daily work?
(RQ2) Do teams design AI systems with traits that align with the traits workers want?

In so doing, we made the following main contributions:

(1) We identified Prolific workers who reported high familiarity with a representative set of 171 tasks spanning 22 occupations drawn from the U.S. Department of Labor's Occupational Information Network (O*NET) (methodology in Section 3.1; task representativeness in Figure 2).
(2) We conducted a scoping review of research on meaningful work (Section 3.2), which informed the construction of survey items for workers (Section 3.3). Building on recent work specifying desirable AI traits across occupations [26], we then developed parallel survey modules: one for workers, eliciting the traits they want AI to possess (Section 3.3), and one for developers, eliciting the traits they intend to design in AI systems (Section 3.4).
(3) We administered the surveys to 202 previously identified workers and to a new sample of 197 developers across 171 tasks in 12 sectors (Section 3.5). We then scaled up their responses by measuring LM agreement with human ratings and, under reasonable agreement, generated task-level annotations for 10,131 tasks across 512 occupations and 19 sectors (Section 3.6). We make these two human-generated datasets publicly available so that future research can explore alternative methods for generating synthetic data from them.
(4) Our first main contribution, addressing RQ1 (Section 4.1; Figure 3), shows that tasks identified as likely to be exposed to AI are more strongly associated by workers with novelty/creativity, positive affect, and autonomy (Figure 4). This challenges the usual narrative that automation mainly targets routine tasks [79]. By contrast, tasks rated as unlikely to be exposed to AI are more often linked by workers to emotional awareness, in-person interaction, relationship building, and social connection (Figure 5).
(5) Our second main contribution, in addressing RQ2 (Section 4.2; Figure 6), reveals clear gaps between workers' preferences and developers' design intentions: workers prefer straightforward systems, whereas developers intend to design polite systems (Figure 6), consistent with reports of LM sycophancy [24], which is a systematic bias of LMs toward agreeing with users' views irrespective of correctness. More generally, developers favor polite, strict, and imaginative systems, whereas workers often describe these design choices as unnecessary friction or rigidity rather than genuine support. By contrast, both groups converge on the need of personalized systems.

Our contributions and findings motivate design principles that aim to preserve meaningful work and better align with worker needs (Section 5), culminating in a five-part research agenda for HCI researchers (Section 6 and Table 2).

2 Related Work

To mirror our research questions in Section 1, we review two strands of literature: (1) which tasks can be exposed to AI (Section 2.1), and (2) how teams design AI to meet workers' needs (Section 2.2; RQ2). We then add a third strand on the meaningfulness of work by conducting a formal scoping review (Section 3.2; Appendix Table 5).

2.1 Tasks exposed to AI

Across field deployments, controlled trials, and exposure analyses, three properties are often associated with AI exposure being likely: (1) the task can be decomposed into explicit steps; (2) the work can be partitioned into sub-tasks under human direction; and (3) outputs can be checked against clear rules, tests, or ground truth. These properties are common in computer-based workflows [28, 32]. In such settings, LMs can accelerate routine components while people supply context and judgment. For example, laboratory studies report faster completion and higher quality in settings where humans set goals and verify results [17, 20, 23, 66, 67].

However, studies also report that few tasks can be automated end-to-end, whereas many can be augmented, often within existing human workflows [28, 32]. Patent-to-task analyses document substantial AI exposure in skilled, non-routine domains (e.g., clinical image review, routing, programming) where professionals nonetheless retain final responsibility [79, 91]. AI exposure is already evident across domains: in software development, AI supports debugging with error traces, proposes refactoring, generates tests, analyzes logs, and prepares code reviews [67]; in professional writing, workers use AI to outline, draft, adjust tone, convert formats, and check citations and style [45, 66]; and in specialist review, clinicians use AI to prepare imaging pre-reads, and draft reports for subsequent human review and sign-off [79].

Even when many tasks could be exposed to AI, empirical studies find that workers often report preferring to retain activities involving judgment, interpersonal interaction, and coordination [9, 18, 28, 32], while accepting greater AI exposure for repetitive digital work [9, 18, 28, 32].

2.2 How teams design AI to meet workers' needs

Prior work suggests that, when tasks feel more meaningful, workers often report preferring to maintain ownership; when they feel less meaningful, workers report being more willing to offload work to AI. Workers judge AI systems typically along two dimensions: warmth (benevolent intent), and competence (ability) [33]. People are more willing to delegate work when they believe the AI system is competent at it [29]. Design choices are reported to shape these beliefs: anthropomorphic features or friendly conversational styles can make AI systems seem warmer and more capable in experimental settings [51, 57]. When people further attribute a 'mind' to AI in such studies, they tend to collaborate with it more, and also blame it more when it makes errors, with these effects amplified by human-like cues [38, 90]. HCI studies show that cues about system expertise, humanness, and fit to the setting influence perceived warmth and competence, with expertise cues most strongly predicting reported trust and use [48, 49]. To complicate matters, system developers often evaluate the trustworthiness of a system differently than system users do [53, 54, 89].

While the warmth–competence framework advances theory, workers rarely judge AI in simple binary terms. Recent evidence suggests people judge whether AI is suitable for a job by the traits the job requires [26]. Dong et al. [26] finds that traits such as fairness, sincerity, warmth, competence, determination, intelligence, tolerance, and imagination were treated as distinct dimensions in assessing whether AI is suitable for a job. Additional studies similarly find that expert cues and context-task fit increase acceptance in workplace settings [49] where examples include interface features that calibrate reliance (e.g., uncertainty cues, targeted explanations), reduce over-reliance, and help people make better decisions [29, 55, 88]. In cooperative tasks, people rely on AI more when it seems warm and competent than when it simply performs well [60]. This suggests that AI system design should distinguish relevant traits based on roles (e.g., fairness for managers, sincerity for clinicians) [60]. In our work, we examine which traits workers want in tasks exposed to AI, and which traits developers design for, revealing role-specific priorities (e.g., fairness for managerial tasks; sincerity for clinical tasks).

Research Gap. Prior work has identified which tasks are exposed to AI and has explored design choices that promote trust and reliability [47]. Yet two important questions remain underexplored. First, we lack understanding of how AI exposure reshapes the meaningfulness of work, and whether tasks exposed to AI feel purposeful to workers or feel merely like bureaucratic busywork. Second, we do not know whether teams design AI systems with the traits that workers actually want. Related work in high-stakes domains such as the judiciary has begun to surface user expectations and requirements for AI tools [82], but little is known about how these expectations translate to everyday work practices across occupations. In the pursuit of productivity and speed, with the introduction of AI, teams may risk changes to autonomy, care for others, excellence, and fairness [43], qualities that prior work links to experiences of meaningful work.

3 Methods

To close this gap, we addressed two research questions (RQs):

(RQ1) Which dimensions of meaningful work do workers associate with tasks exposed to AI in their daily work?
(RQ2) Do teams design AI systems with traits that align with the traits workers want?

To answer these two questions, our methodology followed four steps (Figure 1). First, we selected representative tasks from the O*NET 29.3 database[1], and recruited samples of workers and AI developers on the crowd-sourcing platform of Prolific. We initially identified 171 tasks spanning 22 representative occupations (Section 3.1). By "representative", we mean that the task sample was stratified to approximate the distribution of occupational sectors reported by the U.S. Bureau of Labor Statistics, ending up with modest deviations (Figure 2). Second, we conducted a scoping review synthesizing research on the meaningfulness of work (Section 3.2). Guided by this review, we measured workers' experiences using items that capture the meaningfulness of work (Q1–Q33; Section 3.3). We then measured workers' and developers' views on the design of AI systems using items capturing design traits (Q34–Q45; Section 3.4). Third, we administered the survey to workers and developers under consistent protocols on Prolific (Section 3.5). Fourth, to scale the analysis, we used LMs to simulate workers and developers, generated task-level annotations for 10,131 tasks across 512 occupations and 19 sectors, and validated the model-derived annotations against our human data to assess reliability (Section 3.6).

3.1 Selecting Tasks and Recruiting Workers and Developers

O*NET 29.3 Database. The O*NET database provides standardized information on U.S. sectors[2], occupations and their associated tasks. It organizes 923 occupations into sectors where each occupation is broken down into task statements[3]

[1]https://www.onetcenter.org/database.html accessed July 2025
[2]Sectors refers to O*NET major groups https://www.onetcenter.org/taxonomy/2019/structure.html.
[3]https://www.onetcenter.org/dictionary/29.3/excel/task_statements.html

describing work activities. In total, O*NET contains 18,796 tasks, where each task is classified as *core* or *supplementary*, and annotated with how frequently it's performed (e.g., yearly or less, monthly, weekly, daily, or hourly)[4].

Selecting Tasks Exposed to AI. Our study primarily focuses on workplace tasks likely to be exposed to AI. We follow prior work [80], and apply a multi-step filtering pipeline (Step 1; Figure 1) to identify a representative set of tasks for our study. First, we ensured each task is classified as either *core* or *supplementary* to ensure its relevance to the occupation. Then, we characterized and filtered occupations and tasks by two criteria as determined by GPT-4o annotations in line with [80]: (1) the *occupation* primarily involves computer use and, (2) the *task* can be completed on a computer. Upon manual inspection, we found that GPT-4o occasionally excluded occupations (e.g., nursing, education professionals) that are widely recognized as exposed to AI [70, 74, 75]. To ensure these occupations were represented in our dataset, we manually curated and included a list of 427 occupations that were exempted from these filters, with examples shown in Appendix Table 4. Following these filtering steps, our dataset contained 10,131 tasks spanning 512 occupations.

Recruiting Workers and Selecting Tasks Familiar to Them. We used Prolific to recruit U.S. workers and developers (Appendix Table 7). We applied the "work function" screener to identify participants likely to be familiar with the tasks they evaluated. Both workers and developers were compensated at a rate of $11 per hour. Each O*NET occupation was mapped to one of 21 work functions (Appendix Table 3), yielding 4,473 tasks across 209 occupations. To keep surveys tractable and reduce participant fatigue, we downsampled to at most 10 tasks per occupation using a greedy selection criterion based on task frequency annotations (e.g., daily) and estimated AI exposure [79], resulting in 2,078 tasks across 209 occupations. Our study required participants who were experts in their domains, which made recruitment especially challenging: if participants are not familiar with the tasks they are asked to evaluate, their responses risk being speculative rather than grounded in real-world practice. To target U.S. workers who are experts in their domains, we first ran a preliminary survey which identified workers that: (1) belonged to one of 21 professional work functions on Prolific, (2) passed attention checks, and (3) reported being highly familiar with at least one O*NET task in their occupation. Following this rigorous pre-screening, we recruited 202 workers, excluding those who failed our attention checks. From these responses, we retained tasks rated as "Very familiar" or "Extremely familiar" by at least three participants, producing a final set of 171 tasks across 22 occupations and 12 sectors. This subset provided broad coverage (12 of 22 total sectors in O*NET), while focusing on tasks that were central to the occupation (i.e., core or frequently performed), likely to be exposed to AI, and validated as highly familiar to workers.

Recruiting Developers. For developers, we focused on U.S.-based AI practitioners who met the following screening criteria: weekly AI use (ranging from once a week to multiple times daily), individual contributor or non-manager role, employment in coding, technical writing, or systems administration, and primary function in engineering (e.g., software) or research. We recruited 197 developers who each rated 10 tasks, drawn from the same pool of tasks the workers rated. Each task received ratings from at least three distinct developers. Given the heterogeneity of 'AI developers' on Prolific and the ambiguity of O*NET task descriptions that allows for interpretive variation, we detail the distribution of developers' technical roles, AI usage patterns, and work functions are shown in Appendix Figure 7 to better characterize their expertise. The majority of developers in our sample hold software, data, IT infrastructure, or ML/AI engineering roles. Developers also reported extensive use of AI tools in their workflow, including LLM-based

[4]https://www.onetcenter.org/dictionary/29.3/excel/task_ratings.html

assistants, code-generation tools, data-analysis systems, and ML model development, indicating active engagement with contemporary AI technologies. Work-function distributions further show representation across engineering, research, analytics, and operations roles.

3.2 Scoping Review on Meaningfulness of Work

After selecting tasks and recruiting workers, we next determined which questions would best capture the extent to which workers perceive their tasks as 'meaningful'. To ground these questions in the literature, we conducted a scoping review following the five-stage framework in [6]:

Step 1. Identifying the research question. The main research question was: *What are the documented, theorized, or studied dimensions of task meaningfulness, symbolic work, impression management, and status signaling in today's workplaces?* This question was broad enough to cover both personal views of meaningfulness, and the social or symbolic factors that can make work performative, strategic, or status-driven.

Step 2. Identifying relevant literature. To ensure cross-disciplinary coverage, we went on Google Scholar and JSTOR, and we used the Boolean search string: `(“meaningful work” OR “task significance” OR “work motivation” OR “symbolic work” OR “impression management” OR “status threat” OR “performative work”) AND (work OR job OR employee OR organization OR organization OR labor OR labor)`. The review included only English-language, peer-reviewed work, with no date restrictions.

Step 3. Selecting the articles. We used the following inclusion rules: articles must discuss task meaningfulness, symbolic or performed work, status signaling, or impression management in a work or organizational setting; both empirical and theoretical work was eligible; and full-text access had to be available. Articles outside of work or organizational settings, and those limited to consumer behavior or marketing without reference to employees or task meaning, were excluded, resulting in 56 articles. After removing duplicates and screening titles and abstracts, 42 remained for full-text review, and we included 21 that met all criteria.

Step 4. Charting the data. We coded each article with a structured form. Key fields included: main constructs (e.g., task significance, performative work); theory used (e.g., Job Characteristics Model, Institutional Theory, Impression Management Theory); measures (e.g., Work and Meaning Inventory); and main findings or arguments.

Step 5. Collating, summarizing, and reporting the results. We reviewed 21 articles across psychology, sociology, anthropology, and ethics on what work means to people and to society (Appendix Table 5). The studies show how people judge their own work, how organizations shape those judgments, and how societies value different kinds of work. Examples include David Graeber's critique of "bullshit jobs" [37], a tested questionnaire for meaningful work [83], and research linking work to identity and status [69]. Key theoretical lenses include job characteristics theory, which emphasizes task significance, task identity, and autonomy [41]; impression management, which explains how workers perform roles for symbolic or strategic ends [14, 35]; institutional theory, which highlights symbolic work and routine, ceremonial task structures [61]; and models of status signaling at work [12, 68].

Guided by our coding, we grouped the articles into two analytic levels that together capture the primary sources of experienced meaningfulness:

Micro level (individual appraisal). Articles that tied meaningful work to how people judged their tasks and roles. Findings included task-level features and personal attitudes such as satisfaction, engagement, motivation, and performance [41, 83].

Macro level (organizations, institutions, and society). Articles that explained how society, fields, and organizations ranked different kinds of work and set scripts and norms that shaped meaning, identity, and claims to self-worthiness [12, 14, 35, 61, 68, 69]. This set also reported cases where tasks lacked recognized value and thus felt meaningless [37].

At the micro level, individuals treated meaningful work as an attitude tied to satisfaction, engagement, motivation, and performance. At the macro level, cultural valuations and organizational scripts shape identity and sense making, establishing the norms and constraints that enabled or limited those appraisals.

Prior work has combined insights from the micro and macro levels to explain how individual experiences are shaped by broader organizational and cultural forces. For example, Carton [19]'s study of the National Aeronautics and Space Administration (NASA) in the 1960s argues that macro-level leader 'sense-giving' can recalibrate micro-level experiences. He illustrates how leaders may reshape how people view their work through mid-level links. Using President Kennedy as an example, the study describes leaders as defining a main aim (advancing science), setting a dated goal ('land a man on the Moon before 1970'), outlining a few key steps (Mercury, then Gemini, then Apollo; later a six-step plan sometimes called the 'ladder to the Moon'), and using clear language that ties the goal to shared values such as knowledge and peace. In Carton's account, this plan made the ultimate goal feel more attainable, gave staff clearer stepping stones, clarified their perceived role in the process, and was associated with staff describing daily tasks in mission terms ('putting a man on the Moon', even 'advancing science'). Carton interprets these shifts as aligning motivation, engagement, and performance more closely with the organization's purpose. In a complementary line of work, Bailey et al. [10] integrate psychological and sociological perspectives on how people find work meaningful. They draw on first-person accounts from nurses, creative artists, and lawyers: occupations chosen for their clear contrasts in task content, workplace rules, and room for professional choice, and emphasize the connection between individual experiences and broader organizational purpose for understanding how people experience meaning at work.

3.3 Questions about Dimensions of Meaningful Work to Administer to Workers (Q1-33)

We started with two levels: *individual appraisal* (micro), and *organizational, institutional, and societal valuation* (macro). First, to measure valuation beyond the individual, we drafted *Perceived Bullsh*tness* survey items (Q1–Q5; [36]) and *Status Maintenance* survey items (Q11–Q16; [13, 15]). Second, to measure individual appraisal, we drew on *Perceived Value* (Q6–Q10; [40, 84]), the *EPOCH well-being scale* (Q17–Q21; [58]), and *Human Flourishing* (Q22–Q33; [87]). See Step 2 in Figure 1 and Appendix Table 6 for our complete survey items.

Perceived Bullsh*tness [36] (Q1-5). These five questions measure the extent to which participants view their tasks as pointless, bureaucratic, or not contributing to the goals of their organization. Example survey items include: 'I perform this task only to satisfy bureaucracy or appearances' and 'This task does not contribute to the goals of my organization'. These items build on Graeber [36]'s theory of 'bullshit jobs', which introduces the concept of 'bullshit' jobs as roles that are perceived as worthless, even by those performing them. Graeber [36] argues that these roles can contribute to psychological distress and can erode workers' sense of purpose and motivation.

Perceived Value [40, 84] (Q6-10). These questions assess the extent to which workers perceive a task as meaningful or contributing to the success of their organization. This aligns with the three psychological states (e.g. experienced meaningfulness, experienced responsibility for the outcomes of the work, and knowledge of the results) described in Hackman and Oldham [40]'s Job Characteristics Model. Specifically, this outcome is observed, if an individual '*learns* (knowledge of results) that he *personally* (experienced responsibility) has performed well on a task that *he cares about* (experienced meaningfulness)' [40]. Our survey items reflect this framework by assessing whether workers feel they 'receive useful feedback about how well this task is done.' (Q9; knowledge of results), 'has the freedom to decide how to carry out this task' (Q8; experienced responsibility), and 'provides a sense of accomplishment' (Q10; experienced meaningfulness). This framework is further supported by prior work on meaningful work [84], which shows that seeing one's work as contributing to a greater good is associated with higher well-being and job satisfaction (Q6: 'This task is important to the success of my team or organization').

Status Maintenance [13, 15] (Q11-16). These questions assess the extent to which workers continue performing a task to preserve their professional standing, visibility, and perceived competence. Example survey items include 'I feel this task signals to others that I am busy or valuable', and 'I worry that letting go of this task could reduce my influence or visibility'. These items are derived from prior work on impression management motives [15] where employees engage in behaviors intended to influence how others perceive their abilities, dedication, or value to the organization. Furthermore, related work from consumer research [13] emphasizes how lack of leisure time and 'busyness' serve as status symbols, and they show how individuals may continue to pursue low-value tasks that make workers appear 'busy' because these tasks serve as signals of competence and provide visibility even if they contribute little to core performance outcomes.

EPOCH [58] (Q17-21). These questions capture the extent to which the task requires fundamental human capabilities that prior work argues enable workers to excel in areas where AI is less likely to succeed. Drawing on the EPOCH framework [58], we include items reflecting five dimensions that are particularly challenging to expose to AI: (1) empathy and emotional intelligence ('This task requires recognizing and responding appropriately to the emotions of others'), (2) presence, networking and connectedness ('This task benefits significantly from in-person interaction, non-verbal cues, or spontaneous communication'), (3) opinion, judgment and ethics ('This task involves making decisions that require moral reasoning, accountability, or subjective judgment'), (4) creativity and imagination ('This task requires generating novel ideas, approaches, or solutions beyond standard procedures'), and (5) hope, vision and leadership ('This task involves setting direction, motivating others, or showing perseverance toward a long-term goal').

Human Flourishing at Work [87] (Q22-33). These questions were adapted from VanderWeele [87]'s multidimensional framework of work and flourishing. This framework conceptualizes flourishing as encompassing domains beyond immediate job performance, including well-being, purpose, and social connection. We included survey items to capture six domains: (1) Happiness and life satisfaction (e.g., 'How much would this task make you feel satisfied or content with your work?'), (2) Mental and physical health (e.g., 'To what extent would this task support your mental health?'), (3) Meaning and purpose (e.g., 'To what extent would this task feel meaningful or worthwhile?'), (4) Character and virtue (e.g., 'To what extent would this task allow you to act in accordance with your values or integrity?'), (5) Close social relationships (e.g., 'To what extent would this task help you build or strengthen relationships with colleagues or clients?'), and (6) Financial and material stability (e.g., 'To what extent would this task contribute to your sense of job or financial security?').

3.4 Questions about AI Design Choices for AI Exposure to Administer to both Workers and Developers (Q34-45)

Even when workers wish to use AI, prior studies show that AI tools often fail to meet their needs due to limited understanding of which psychological traits workers expect AI systems to exhibit [26, 80]. Adoption and acceptance of AI technologies depend on the extent to which these systems align with user values and expectations [86]. Yet, value alignment is dynamic: values emphasized at the design stage often diverge from those prioritized by users once technologies are deployed in real contexts [46]. Prior work has shown that values essential to effective task performance (e.g., empathy, fairness, creativity) are rarely embedded into the design of AI systems [58]. To investigate this gap, we assessed workers' preferences for how an AI system should behave, if their tasks were exposed to it. Specifically, Questions 34–45 (Step 3 in Figure 1, and Appendix Table 6) asked participants to rate the importance of twelve psychological traits for an AI system to exhibit: four traits that we introduced (Q34–Q37: creativity, empathy, explainability, and openness to challenge) based on the HCI literature [3, 65, 81], and eight traits taken from Dong et al. [26] (Q38–Q45: fair, warm, sincere, tolerant, competent, determined, intelligent, and imaginative).

3.5 Administering Questions to Workers and Developers

We administered two complementary surveys: one to workers, and the other to developers. The worker survey was designed to capture perceptions of workplace tasks and preferences for AI system behavior and consisted of Questions 1–45. The developer survey was designed to only capture preferences for AI system behavior and, thus, consisted of Questions 34–45 (Appendix Table 6). Each item was rated on a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree). Both surveys concluded with the Human Agency Scale (HAS) [80] (e.g., Q48), which assessed desired levels of human–AI collaboration. The scale ranges across varying degrees of human involvement: from AI agent drives task completion (HAS H1–H2), to equal partnership (HAS H3), to human drives task completion (HAS H4–H5). In the survey, workers were asked: '*If AI were to assist in this task, how much of your collaboration would be needed to complete this task effectively*'? Response options reflected the five HAS categories, allowing us to examine worker preferences for human intervention and to contrast these with the intervention priorities of developers.

3.6 Scaling and Validating Responses with Language Models

To enable larger-scale analysis, we evaluate whether LMs can act as annotation assistants to simulate the distribution of worker and developer responses. Recent work has demonstrated the promising potential of LMs as annotation assistants in social science settings [71, 72, 78] where LMs can approximate human judgments in large scale surveys [5].

We used in-context learning with GPT-4o, applying chain-of-thought prompting [92] to adopt the persona of either a worker or a developer for a given occupation. This procedure is briefly described in Step 4 of Figure 1, and the prompts are fully detailed in Appendix Tables 9 and 10. Our approach aligns with prior work on LM-as-an-Expert prompting [44, 62, 93], which has been validated as a method for eliciting domain-specific expertise from LMs. We then assessed the external validity of our findings based on LM-generated annotations (i.e., whether the patterns we report are consistent with and generalize to human judgments) using three complementary strategies:

(1) **Testing whether incorporating the LM as an additional annotator improved inter-rater agreement.** To evaluate the reliability of human annotations, we first assessed the reliability using only human annotators ('experts') by calculating intra-class correlation coefficients (ICC) and mean absolute differences (MAD) at the task level for items with at least three expert ratings. For the worker survey, experts alone achieved a mean ICC of 0.634 (moderate-to-good agreement; 95% CI = [0.602, 0.664]). Adding the LM as an additional annotator increased the mean ICC to 0.742 (good agreement; 95% CI = [0.722, 0.760]), a statistically significant improvement of +0.108 (95% CI = [0.093, 0.124]). For developer responses, experts alone achieved a mean ICC of 0.629 (moderate agreement; 95% CI = [0.578, 0.676]). Adding the LM increased the mean ICC to 0.673 (moderate agreement; 95% CI = [0.627, 0.713]), representing a statistically significant improvement of +0.044 (95% CI = [0.032, 0.058]). For workers, LM ratings differed from experts by about 1.10 Likert points (normalized MAD = 0.276), only slightly higher than expert–expert disagreement (0.255). For developers, LM–expert differences averaged 1.31 points (normalized MAD = 0.328), which was not statistically significantly different from expert–expert disagreement (0.324). To further assess whether ICC improvements reflected genuine alignment rather than artifacts of increased rating stability, we conducted a robustness analysis by comparing the real LM's contribution to that of a randomized version that preserved its overall rating distribution but no longer reflected task-level correspondence with human ratings, by randomly shuffling its existing ratings. We then repeatedly added this randomized LM to the human ratings and recomputed ICC across 1,000 bootstrap samples. Adding the randomized LM yielded modest ICC increases (+0.079 for workers; +0.016 for developers), showing that some improvement comes from variance stabilization. However, these increases were significantly smaller than those from the real LM (+0.108 for workers; +0.044 for developers; $p < .0001$). This indicates that the ICC gains arise from substantive alignment between the LM and human annotators, not statistical artifacts.

(2) **Comparing annotation distributions between LM and human raters, supplemented with qualitative analyses.** Across the dimensions of meaningful work, LM ratings closely tracked human ratings, with only minor differences in distributional shape (Appendix Figure 8). In addition to computing global reliability indices, we examined the percentage of large discrepancies, defined as cases where human and LM ratings differed by two or more points on the 5-point Likert scale. Whereas a one-point difference can be attributed to normal subjectivity, a gap of two points or more represents a substantive divergence in interpretation [21]. Identifying such items provides a diagnostic lens on the alignment of LM outputs with human ratings, revealing systematic areas of disagreement that global coefficients such as ICC may obscure. We find that only 0.87% of tasks exhibit such significant discrepancies. Across dimensions, the *EPOCH* questions showed the highest divergence at 2.92%, followed by *AI Design Choices* questions at 1.75%. All remaining dimensions (aside from a few AI-trait questions) had negligible discrepancy rates, each below 1%. In these few cases, we observed that both LM and human interpretations were reasonable when viewed from different contextual perspectives (Appendix Table 8). For example, the LM perceived the task '*review, classify, and record survey data in preparation for computer analysis*' as a routine procedure with minimal emotional demands, whereas a human annotator emphasized its creative and moral judgment aspects. Crucially, these divergences are best understood as subjective differences rather than systematic bias, and, as we show in our third strategy, they do not affect the substantive results when analyses are replicated using only human ratings.

(3) **Replicating the main analyses using only human ratings.** We replicated the core analyses that we will present in Section 4 using only human ratings. The replication yielded results that were highly consistent with those based on LM-generated ratings across both research questions, thereby providing a direct test of

robustness. For RQ1 (Appendix Figure 9), humans and LMs agreed on 6 of the 7 meaningful-work dimensions that differentiated tasks likely to be exposed to AI from those unlikely to be exposed. The only exception was '*requires novel ideas or creativity*'. For RQ2, we conducted a head-to-head comparison between human ratings and LM-simulated ratings on our subset of 171 tasks. As shown in Appendix Tables 17 and 18, the pattern of worker–developer misalignment produced by human raters closely tracks the pattern produced by LM-simulated raters, with consistent rankings across high-, mixed-, and aligned traits. The strongest pattern was workers emphasizing straightforward traits, and developers emphasizing politeness traits, which was reproduced almost exactly in human-only ratings. Divergences appeared only in secondary dimensions (e.g., 'handle complex *vs.* routine work', 'precise *vs.* simple'). Manual review suggests this reflects sectoral biases introduced by the smaller, occupation-concentrated human sub-sample rather than any substantive difference in trait interpretation. Importantly, none of these divergences alter the main inferences: the central dimensions of meaningful work, and the dominant sources of worker–developer disagreement are consistent across LM and human annotations.

These results indicate that, in our setting, GPT-4o can serve as a reliable additional annotator for both worker and developer perspectives without reducing inter-rater reliability. We therefore used LM-generated ratings to scale our analysis to the full set of 10K O*NET tasks, spanning all 19 occupational sectors. The LM-annotated dataset provides broader coverage than the worker and developer surveys, with task distributions more closely aligned with the Bureau of Labor Statistics (Figure 2), supporting large-scale analysis of workforce patterns.

4 Results

Before presenting the results in depth, we provide a brief overview with references to the sections where each finding is discussed. In summary, we observed that:

(1) *Creative and high-agency tasks are more exposed.* Across sectors, tasks in the likely-to-be-exposed group tend to emphasize creativity, positive affect, and autonomy (Section 4.1). Sectors with higher scores on these traits include Arts, Architecture & Engineering, Computer & Mathematics, and Life, Physical, & Social Science (Figure 4). This pattern contrasts with the familiar narrative that automation will primarily absorb routine tasks, freeing workers to concentrate on higher-value activities such as strategy and design. Our results suggest a more complex trajectory. Generative systems are already used to draft text, suggest layouts, start campaigns, run simulations, and infer likely emotional responses. In our data, these uses are associated with tasks that workers describe as meaningful because they reflect taste, judgment, and authorship. Rather than being confined to low-level chores, AI systems are increasingly entangled with how people add value to work: from generating first ideas to editing, selecting, and retaining accountability.

(2) *Social and face-to-face tasks are less exposed.* Tasks rated as not likely to be exposed depend more on emotion, in-person contact, social ties, and support (Section 4.1). These cluster in Community & Social Service, Education, Healthcare, and Sales (Figure 5).

(3) *Worker–Developer misalignment.* We finally found systematic misalignment between how workers want AI systems to behave and how developers intend to design them (Section 4.2). Developers tend to emphasize politeness, strictness, and imagination, especially in high-stakes or highly structured domains whereas workers often describe such traits as sources of delay or rigidity rather than support.

Fig. 2. Proportions of tasks across occupational sectors in three datasets: Prolific (n=171), LM-annotated (n=10K), and U.S. labor statistics (n=18K). The Prolific sample covers 12 sectors (based on available occupations), the LM-annotated dataset covers 19 sectors, and the U.S. Bureau of Labor Statistics dataset covers all sectors. The distributions of Prolific and LM-annotated tasks are broadly similar across sectors, indicating that LM annotations capture sector patterns consistent with worker data.

4.1 Which dimensions of meaningful work do workers associate with tasks exposed to AI? (RQ1)

Our goal in RQ1 is to examine whether tasks that are more likely to be exposed to AI differ systematically in the significance they hold for workers. Do they call for novel ideas? Are they associated with feelings of agency? Do workers link them to relationship building or emotional awareness?

To test this, we partitioned our tasks into likely-to-be-exposed *vs.* not-likely-to-be-exposed groups, and restrict the sample to computer-based occupations [80]. We then fit item-wise linear mixed-effects models on the 33 worker survey items (Q1–Q33), using exposure likelihood as a fixed effect and random intercepts for sector and occupation. We first identify which dimensions of meaningful work are disproportionately exposed to AI, then examine sector-level patterns, and provide task examples that align with statistically significant dimensions of meaningful work.

Likely-to-be-exposed *vs.* Not-likely-to-be-exposed Tasks. We divided our 10,131 tasks (LM-annotated) into two groups: *likely-to-be-exposed vs. not-likely-to-be-exposed.* We used AII [79] to estimate the likelihood that workplace tasks will be exposed to AI. Following Shao et al. [80], we further restricted both groups to occupations and tasks that are primarily performed on a computer according to O*NET, resulting in 3,179 tasks across 426 occupations and 19 sectors in the likely-to-be-exposed group, and 2,349 tasks across 381 occupations and 19 sectors in the not-likely-to-be-exposed group.

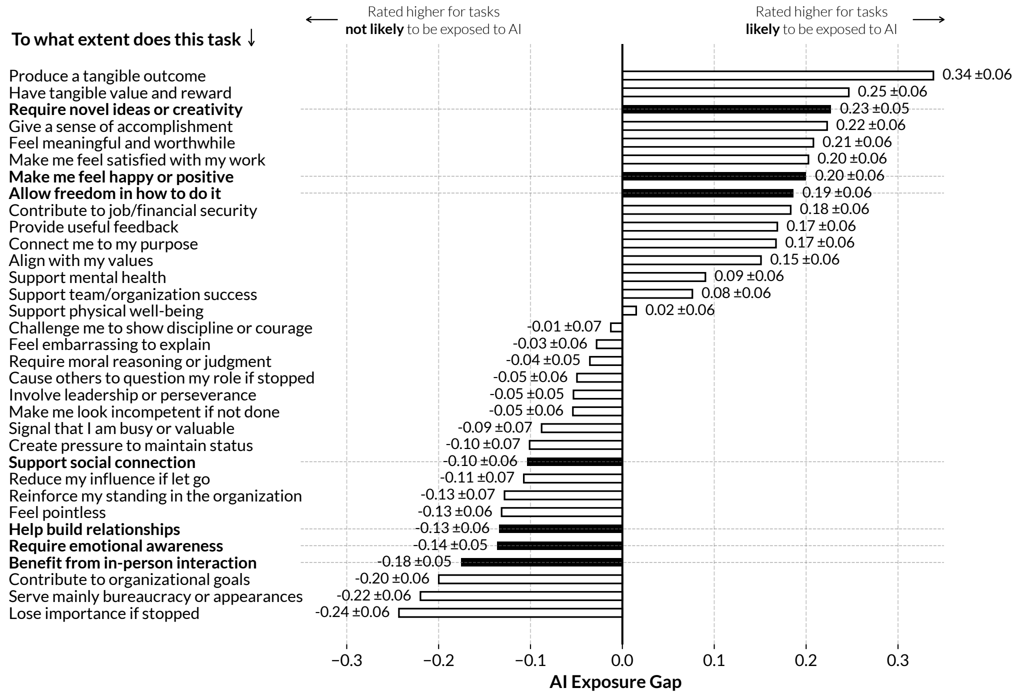


Fig. 3. AI Exposure Gap by dimensions of meaningful work (rows). The higher the gap, the more strongly that dimension is associated with tasks likely to be exposed AI. This gap is computed as the difference of how important a dimension is between two groups of tasks: those that seem more likely to be exposed to AI and those less likely. We estimate the gaps and 95% confidence intervals with mixed-effects models. Bold names and corresponding black bars indicate differences that are statistically significant. Tasks rated likely to be exposed tend to involve novelty, creativity, happiness, and freedom in how workers do them. Tasks rated not likely tend to involve emotional awareness, in-person interaction, building relationships, and supporting social connection.

Linear Mixed-Effects Regression Model. We estimate whether tasks likely to be exposed to AI systematically differ in their meaningfulness to workers as compared to tasks not likely to be exposed. For each task t, we computed the importance of each dimension d of meaningful work (e.g., requiring novel ideas or creativity, help build relationships). To estimate whether each dimension was more or less important in tasks likely to be exposed than in tasks not likely to be exposed, we fit a linear mixed-effects regression of the form:

$$z(y_{t,d}) = \beta_0 + \beta_1 \cdot \text{AIExposure}_t + u_s + u_{o(s)} + \epsilon_{t,d},$$

where $z(y_{t,d})$ is the z-score of the rating $y_{t,d}$ a worker gave to the importance of dimension of meaningful work d for task t; AIExposure_t is a binary variable equal to 1, if task t is likely-to-be-exposed, or equal to 0, if task t is not-likely-to-be-exposed; β_0 is the fixed effect intercept, representing the baseline importance of dimension d for not-likely-to-be-exposed tasks ($\text{AIExposure}_t = 0$); β_1 is the fixed effect of AI exposure, estimating the mean difference in worker ratings between likely-to-be-exposed and not-likely-to-be-exposed tasks. Given that tasks are nested within occupations, which, in turn, are nested within sectors, to account for varying baselines within sectors and occupations, we also included a random intercept u_s for sector s, and random intercept $u_{o(s)}$ for occupations nested within sector s,

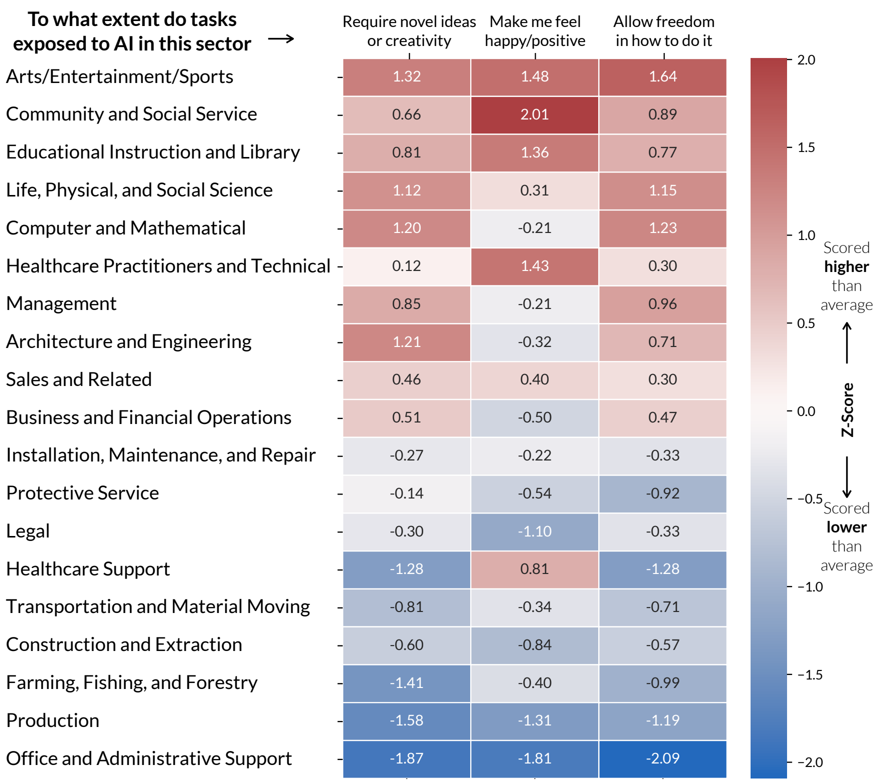


Fig. 4. Association of tasks exposed to AI in each of the sectors (in the *rows*) with subset of three dimensions of meaningful work (creativity, positive affect, and autonomy in the *columns*). Sectors are sorted by the average z-score across the three dimensions. Creative and socially-oriented sectors (arts, community service, education, life sciences) are associated with tasks exposed to AI that emphasize novelty, positivity, and freedom. In contrast, routine and manual sectors (office support, production, farming) score much lower.

followed by $\epsilon_{t,d}$, which is the residual error term for task t on dimension d. Models were estimated using maximum likelihood (restricted maximum likelihood, REML, disabled). For each dimension, we report β_1 (the difference in worker ratings for likely-to-be-exposed *vs.* not-likely-to-be-exposed tasks), its standard error, and 95% confidence intervals. To address multiple comparisons across our survey items, we applied the Benjamini–Hochberg False Discovery Rate (FDR) correction. We defined the fixed effect as statistically significant, if two conditions were met: (1) the FDR-adjusted $p < 0.05$, and (2) the effect size exceeded $\Delta \geq 0.1$ Likert points. While a threshold of 0.1 may appear small, our results are reported in aggregate across all sectors; when disaggregated at the sector level, we show that differences are often substantially larger. The latter serves as a threshold for practical significance: on a 5-point scale, a 0.1 shift represents a small but interpretable change in perceived task characteristics, ensuring that we highlight effects that are not only statistically detectable but also meaningful in practice.

The mixed-effects estimates and FDR-adjusted tests (see Appendix Table 11) yield three overarching findings about how AI exposure shapes the perceived importance of meaningful-work dimensions, and how these effects distribute across sectors:

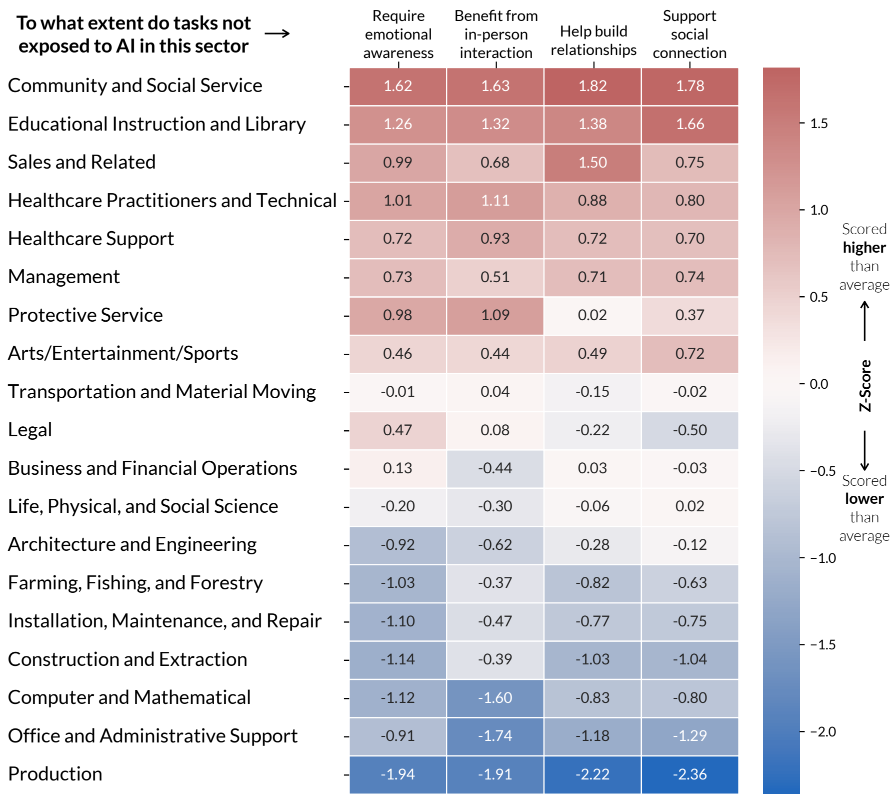


Fig. 5. Association of tasks not exposed to AI in each of the sectors (in the *rows*) with subset of four dimensions of meaningful work (emotional awareness, in-person interaction, relationship building, and social connections in the *columns*). Sectors are sorted by the average z-score across the subset of four dimensions. Human-facing sectors such as community and social service, education, and healthcare consider their tasks not exposed to AI to emphasize emotional awareness, in-person interaction, and social connection. In contrast, technical and routine sectors (e.g., production, office support, computer and mathematical) score far lower, indicating that workers in these areas view tasks not exposed to AI as less socially or emotionally significant.

(1) *The tasks most exposed to AI involve creativity and high levels of individual agency, while tasks that rely on empathy, relationship-building, or in-person presence appear less exposed.* Across a subset of seven significant dimensions of "meaningful work", those most exposed to AI emphasize novelty and creativity, personal agency, and the capacity to elicit positive emotions, whereas tasks less exposed emphasize social connection, relationship building, emotional attunement, and in-person interaction (Figure 3, and Appendix Table 11). Random-slope mixed-effects models allowing AI exposure to vary by sector showed largely consistent effects across sectors, with notable heterogeneity for visible/tangible outcomes, emotional awareness, in-person interaction, and physical well-being (likelihood-ratio tests, $p_{\text{LRT}} < 0.05$).

(2) *Tasks highly-exposed to AI cluster in creative, technical, and scientific domains, where AI systems increasingly support ideation.* The greatest exposure appears in the Arts, Architecture & Engineering, Computer & Mathematics, and the Life, Physical, & Social Sciences (Figure 4). Illustratively, an Art Director "formulating basic layout design" and an Actuary "constructing probability tables for natural disasters" reflect first passes that models

now credibly generate, with humans then refining the output (Appendix Table 12). These sectoral patterns align with cluster analyses of high-importance tasks (Appendix Tables 13-15). Tasks likely to be exposed to AI that evoke positive emotions are concentrated in Arts, Entertainment, Sports, & Media; Community & Social Services; and Healthcare (Figure 4).

(3) *Tasks that remain less exposed to AI are those that rely on relationships and sensitivity to context, with their value derived from human attention and judgment.* In Education, Sales, Community & Social Services, and Healthcare, core activities (e.g., "counseling students through intertwined academic and personal issues", "presenting offers while preserving relationships") depend on real-time, co-constructed meaning, and nuanced perception that resists codification (Figure 5, and Appendix Table 12). Consistent with this pattern, clusters emphasizing emotional awareness and relationship building are predominantly human-centered (Appendix Table 16).

4.2 Do teams design AI systems with traits that align with the traits workers want? (RQ2)

For RQ2 (Q34–Q45), we were interested not in dimensions of meaningful work, but in *AI traits*. That is, we asked workers and developers which traits an AI system should have. Each item defined a trait along a spectrum (e.g., Q39: "*Should the AI show warmth and care, or remain neutral and businesslike?*"), and participants rated their preference on a 1–5 scale. Worker responses indicated which traits an AI system should have when their tasks are exposed to it; developer responses indicated how practitioners would design such a system. Should the system be straightforward or polite? Tolerant or strict? Practical or imaginative? Flexible or determined?

To measure misalignment, we subtracted the rating workers assign to trait q for an AI system exposed to task t (workers rating$_{t,q}$) from the rating developers assign to the same trait for the same task (developers rating$_{t,q}$):

$$\Delta_{t,q} = \text{workers rating}_{\text{t},q} - \text{developers rating}_{t,q}.$$

The magnitude $|\Delta_{t,q}|$ measures the size of the misalignment, while the sign indicates direction. For example, a positive $\Delta_{t,q}$ on Q39 indicates that workers preferred more warmth and care than developers, who leaned toward neutrality. We averaged task-level differences for each sector giving each task equal weight, giving us an average misalignment score per sector. To then test whether the differences between worker and developer ratings were statistically significant, we conducted a *two-sided t-test* of the null hypothesis that the mean task-level difference was zero for a given trait and sector. To account for multiple comparisons, p-values were adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) procedure. A sector was labeled as significantly misaligned on a trait, if two conditions were met: (1) the FDR-adjusted $p < 0.05$; and (2) the absolute mean difference ($\frac{1}{N}\sum_{t=1}^{N}|\Delta_{t,q}|$; where N is the number of tasks in the sector) exceeded a threshold of 0.5 Likert points. We use 0.5 as a conservative threshold to focus on practically meaningful differences: unlike RQ1, which examined fine-grained within-task effects (where smaller shifts of 0.1 were informative), RQ2 compares worker–developer ratings aggregated across major occupational groups, where only larger gaps are more informative.

Most and Least Misaligned Traits. To analyze worker–developer misalignment, we ranked AI traits by their average misalignment scores across sectors where worker/developer differences were statistically significant for a given sector (FDR < 0.05, $\frac{1}{N}\sum_{t=1}^{N}|\Delta_{t,q}| \geq 0.5$). Table 1 summarizes the classification of traits from most to least misaligned, while Figure 6 illustrates example occupations and tasks within each category. Sector-level misalignment scores for individual traits are reported in Appendix Tables 25-30. To summarize our results, we see that, across sectors (Figure 6),

Table 1. Worker–developer misalignment by AI traits. Misalignment is defined as the average absolute difference between worker and developer ratings (Q34–Q45) of the traits they believe AI systems should possess. Differences ($\Delta_{t,q}$) are calculated as worker minus developer ratings for a given task t for a trait q, with the magnitude ($|\Delta_{t,q}|$) reflecting the size of the misalignment. Reported values are aggregated over sectors and traits are grouped into high, mixed, or aligned categories based on percentile thresholds of average absolute misalignment. # Sig. Sectors refers to the number of sectors that had statistically significant *differences* in ratings between workers and developers. The largest gaps appear for Straightforward *vs.* Polite, Tolerant/Open-minded *vs.* Strict, Practical *vs.* Imaginative, and Flexible *vs.* Determined, whereas traits such as Generalized *vs.* Personalized, Simple *vs.* Comprehensive, and Business-like *vs.* Warm/caring show little to no misalignment.

Trait	**# Sig. Sectors**	$\Sigma\lvert\Delta_{t,q}\rvert$	$\mu\lvert\Delta_{t,q}\rvert$
High misalignment			
(Q40) Straightforward *vs.* Polite	16	25.874	1.617
(Q41) Tolerant/Open-minded *vs.* Strict	5	4.449	0.890
(Q45) Practical *vs.* Imaginative	2	1.690	0.845
(Q43) Flexible *vs.* Determined	9	6.654	0.739
Mixed misalignment			
(Q34) Handle complex *vs.* Routine work	8	5.797	0.725
(Q35) Address emotions *vs.* Apathetic	4	2.849	0.712
(Q36) Explainable *vs.* Fast/automatic	16	10.940	0.684
(Q42) Precise *vs.* Simple	2	1.168	0.584
Aligned			
(Q37) Definitive *vs.* Open to challenge	3	1.713	0.571
(Q39) Business-like *vs.* Warm/caring	1	0.500	0.500
(Q38) Generalized *vs.* Personalized	0	0.000	0.000
(Q44) Simple *vs.* Insightful/Comprehensive	0	0.000	0.000

workers consistently favored straightforward systems; developers preferred polite ones. Workers wanted tolerance; whereas developers leaned towards more strict systems. Workers asked for practical systems; developers opted for more imaginative systems. Workers liked flexibility; developers nudged toward more determined systems. Where both groups aligned was telling: they valued deep understanding, personalization, and openness to challenge. That is, neither group preferred a generic system that appeared helpful but functioned as an unquestionable authority.

Also, to surface broad patterns of misalignment, for each trait, we identified tasks in the extreme percentiles of misalignment (top 99^{th}, and bottom 1^{st}), clustered these tasks using MPNet[5] embeddings and K-Means clustering, and labeled the resulting clusters (Appendix Tables 19-24). To interpret these aggregate results, we distilled three recurring, salient design tensions:

(1) *When the stakes are high, workers often treat 'politeness' as a delay rather than a feature.* The disagreements varied substantially across sectors. The politeness divide, in particular, was most pronounced in Production, Architecture & Engineering, and the Life, Physical, & Social Sciences, fields where vagueness can result in wasted materials, structural failures, or flawed data. When we clustered the most misaligned tasks on politeness (Appendix Table 22). This clustering highlighted the parts of the economy that demand exacting judgment: quality control, technical design, oversight, and coordination. The corresponding tasks resemble the day-to-day

[5]https://huggingface.co/sentence-transformers/all-mpnet-base-v2

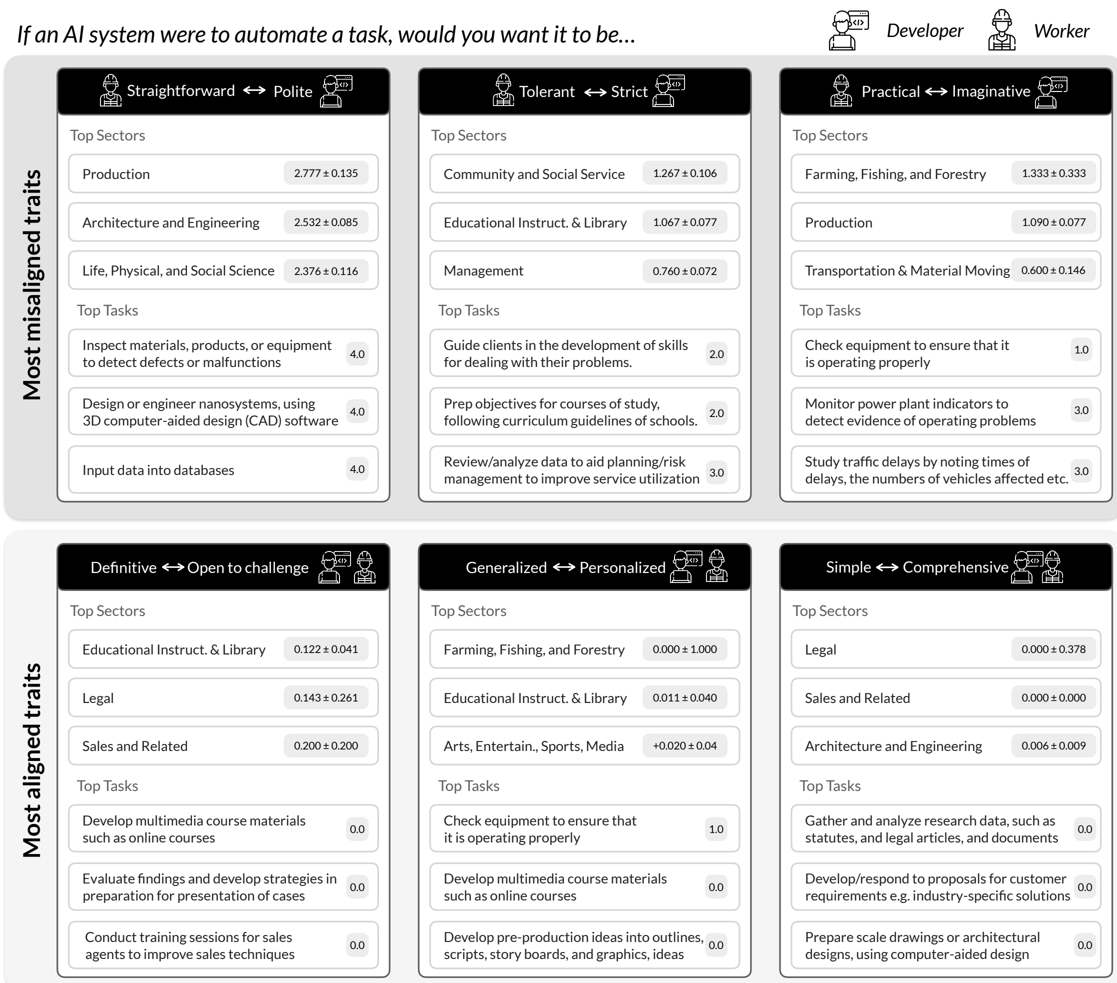


Fig. 6. The three most misaligned AI traits (top row) and the three most aligned traits (bottom row). Larger values indicate greater disagreement between workers and developers. Scores are the absolute differences between workers' ratings of how much they want an AI system to exhibit each trait and developers' ratings of how much they intend to design that trait into an AI system. Icons show which trait direction each group prefers (e.g., workers wish straightforward systems, while developers set out to design polite systems). Top contributing sectors and example tasks are listed for each trait.

activities of highly skilled workers such as planning stress tests, analyzing medical procedures to forecast outcomes, or coordinating complex projects.

(2) *Workers wanted AI systems that are flexible; developers tended to value strictness.* Along the tolerant–strict dimension, divergences were most pronounced in Community & Social Service ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = +1.27$), Education ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = +1.07$), and Management ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = +0.76$) (Appendix Table 28; N is the number of tasks in a sector). Although these settings might be presumed to benefit from greater structure, in practice, "strict" software is often experienced as rigid rule-based constraints that can limit practitioner judgment. Those

domains are indeed characterized by frequent exceptions, and context-sensitive decision-making: for example, accommodating late coursework without disproportionate penalty, or processing intake information that does not conform to standardized fields. In Management & Education, as shown in Appendix Table 23, task clusters include process improvement, monitoring and planning, and budget or risk management, where developers emphasize stricter standards and oversight, while workers prefer tolerance and flexibility.

(3) *Creativity is not always a virtue.* Along the practical–imaginative dimension, developers tended to favor more imaginative systems, whereas workers prioritized pragmatism. The largest divergences were in Farming, Fishing, & Forestry ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = -1.33$), Production ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = -1.09$), and Transportation & Material Moving ($\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q} = -0.60$). These are domains characterized by highly structured, routine workflows (Appendix Table 30). Workers engaging in equipment checks, plant monitoring, and traffic analysis preferred systems that detect anomalies, enforce compliance, and adhere closely to task constraints rather than tools oriented toward open-ended ideation. Task-level examples are illustrative (Appendix Table 24): checking equipment to ensure proper operation (Farming); monitoring power-plant indicators to detect operating problems (Production); and studying traffic delays by recording times and vehicle counts (Transportation). Consistent with this pattern, clustering results (Appendix Table 21) indicate that the most misaligned tasks are concentrated in routine technical operations: equipment monitoring, compliance, and other highly structured activities that emphasize accuracy, continuous monitoring, and rule adherence.

5 Discussion

In Section 5.1, we relate our findings to prior work and summarize our main empirical contributions. Section 5.2 then draws out implications for the design and governance of AI systems, and Section 5.3 concludes by discussing limitations of our study and directions for future research.

5.1 Relation to Prior Work and Overview of Findings

Relation to prior work and novel contributions. Prior work has found that, in many settings, augmentation is more common than end-to-end automation [16, 43, 80]. Our task-level analysis introduces a worker-centered perspective by examining how exposure relates to how work feels to workers, and by identifying the aspects they prefer to handle through in-person, social interaction. This complements exposure inventories by bringing in how tasks are experienced by workers, and by highlighting potential implications for design [16, 43, 80]. More specifically, we make three main contributions:

(1) *A task-level link between exposure and meaning.* We analyze how AI exposure co-varies with dimensions of meaningful work and observe a clear pattern: exposure is higher for tasks involving new ideas, positive feelings, and freedom, and lower for tasks that rely on emotional attunement and in-person relationships (Figures 3, 4, and 5) [43, 80].

(2) *A worker–developer trait map for AI exposure.* We quantify where workers and developers differ on AI traits by sector and task. We find stable agreement on personalization and deep comprehension/insightfulness, and large gaps on "straightforward *vs.* polite" (Table 1, Figure 6, and Appendix Figure 10) [79].

(3) *A scalable rating process with checks.* We pair human ratings with LM-assisted ratings to cover the largest set of tasks present in the literature, while ensuring validity. In our data, adding the LM as an additional annotator

was associated with higher inter-rater agreement for both worker and developer instruments, and we document where LM and human views diverge (Table 8, and Appendix Figure 8) [94].

Workers and developers differ in specific AI traits. We briefly compare where they differ and where they agree, and then suggest a broader design implication:

Where they differ most. Workers want AI that is straightforward, tolerant and open-minded, practical, and flexible. Developers plan for AI that is more polite, strict, imaginative, and determined. The biggest gaps appear for straightforward *vs.* polite, tolerant *vs.* strict, practical *vs.* imaginative, and flexible *vs.* determined (Figure 6 and Table 1). Sector gaps for straightforward *vs.* polite are largest in Production, Architecture & Engineering, and Life, Physical, & Social Sciences, with smaller gaps in Education, and Community & Social Service (Figure 10).

Where they agree. Both groups favor deep understanding/comprehension, personalization, and openness to challenge, with little to no systematic gap (Figure 6 and Table 1).

Design Implication. Our results suggest that AI development teams should compare planned trait choices to worker preferences for the target task and sector. In domains with technical judgment and oversight, large gaps on straightforward *vs.* polite, and tolerant *vs.* strict call for careful defaults, clear settings, and a broader set of design choices (Section 6).

5.2 Implications

We translate our results into three steps for deployment: design the interaction to protect meaning, measure the outcomes that matter, and tune defaults by sector and task.

Design the interaction to protect meaning. An interface that makes assistance easy to accept, edit, and credit (without reducing the worker's role) is consistent with our findings. Control stays at the task level when the system exposes levels of help at the sub-step (by, e.g., suggesting, drafting, or executing), and defaults to reversible suggestions that require explicit confirmation before applying changes. Clarity about edits and credit comes from showing sources and a simple revision history, which makes authorship and changes visible in the final output. Tone should match the work: when a straightforward style is preferred, a plain default with an optional tone control should be suitable (Figures 6 and 10).

Measure what the design seeks to preserve. We recommend measuring, during deployment, whether the design preserves human control, supports learning, and assigns credit fairly, without sacrificing speed or quality. Short, task-linked metrics should capture latency and output quality, while logs record the assistance level (suggest, draft, execute) and the final decision-maker. Regular reviews of these logs can help teams monitor whether deployments remain aligned with meaningful assistance rather than drift toward replacement.

Set sector-aware defaults. We recommend setting sector-specific defaults that reflect the task differences found in our results (Figures 6 and 10). For technical oversight and design, defaults should be simple, practical, and adjustable. For care and education, defaults should be warm and personal while keeping limits explicit. More generally, in contexts where emotional awareness and in-person interaction are essential, our results point toward using AI primarily for background tasks (e.g., preparing briefs, summarizing records, and flagging anomalies), while reserving protected time for direct human engagement.

Our findings suggest an *interaction-as-policy* lens: decisions about who clicks, who decides, who sees what, and what the AI system makes easy or hard are not just UX choices but *governance* choices embedded in the interface. This builds on HCI and STS work that treats infrastructures, defaults, and algorithms as forms of governance (e.g., code as law, scripts that configure users, and algorithmic management in workplaces) [2, 46, 73]. This lens helps organize four observed phenomena:

(1) *Situated manners.* Situated manners recast politeness in HCI as a context-sensitive control, not a universal style. "Frictionless and chatty" helps in consumer chat, but in operational work, where risk, time pressure, irreversibility, uncertainty, and physical coupling matter, verbosity and small talk distract. The system should default to short, clear outputs and strategic silence. Examples from Section 4: disaster-risk tables should present the number, the limit, and the next action; for purchase offers, the system should surface price, terms, and deadline first; in counseling, the system should adopt a warmer tone and avoid imperatives. In high-risk contexts, the ethical choice may well be the blunt one, and the useful thing to say may be short.

(2) *Liability anxiety.* Our results are consistent with the idea that liability concerns may contribute to preferences for strictness. In our "Regulatory requests for information" cluster (Legal, Sales in Section 4), stricter configurations may have reduced wrongful disclosures and minimized regulatory or contractual exposure, but may have increased missed statutory windows and incomplete filings despite available data.

(3) *Romanticized creativity.* Developers often romanticize creativity as "frontier intelligence", but our findings suggest that creativity is most helpful when tightly sequenced and scoped. In field troubleshooting, operators want diagnosis first (fault codes, likely failure chains, next safe test). In regulatory responses, creative paraphrase undermines audit-readiness. In customer support, agents prefer policy-aligned drafts with required fields pre-filled over lively copy that risks unauthorized promises.

(4) *Power structures.* Our analysis treats interaction design as a form of policy, but it is important to ask *who sets that policy.* Our results show that tasks with high levels of creativity, freedom, and happiness are especially likely to be exposed to AI. In practice, managers, vendors, and technical teams (not workers) usually decide which tasks to expose to AI, how far to automate them, and how to evaluate success, often under pressure to improve efficiency [50]. This pattern parallels familiar power asymmetries from work on algorithmic management, where data-driven systems restructure tasks, monitor performance, and allocate rewards in ways that can increase organizational control over workers [46, 73]. In our setting, interaction patterns (such as one-click automation and suggestion-on-demand) shape whether AI replaces or supports the parts of the job that workers report finding most meaningful. Managers, vendors, and technical teams make these decisions before workers can consent or take part, especially in roles with low bargaining power. Therefore, beyond designing interfaces that let individuals retain agency within a task, organizations should give workers some say over which task clusters are candidates for automation. When such mechanisms for voice are absent, AI exposure risks deepening existing power imbalances. Prior work on algorithmic management and worker-centered AI points to several mitigations: involving workers and their representatives directly in technology decision-making (through, e.g., formal consultation and collective bargaining over AI deployments), creating worker or union "technology representatives" with access to information about how systems allocate and evaluate work, and requiring transparency tools or reports that surface the indicators workers need to understand and contest algorithmic decisions [2, 22, 30, 73].

5.3 Limitations

This work has several broad limitations:

(1) *Representativeness of tasks and participants.* We have focused on U.S. occupational tasks. Results may differ in other regions or settings with other norms or tools [16]. Our worker and developer samples were recruited through Prolific, where participants tend to be more technologically literate. As a result, perspectives from low-wage, non-digital, or less AI-exposed occupations may be underrepresented. Although our pre-screening procedure ensured that participants were highly familiar with the tasks they evaluated, future work should expand to other recruitment channels (e.g., industry partnerships, unions, vocational training programs) to capture a wider diversity of workplace contexts and skill levels. Although the distribution of technical roles, AI usage, and work functions in our sample indicates that developers were actively engaged in building or maintaining AI-driven systems across diverse sectors (Figure 7), we acknowledge that our samples may not fully capture the breadth of industry roles.

(2) *Modeling assumptions.* Our mixed-effects regression models assume linear relationships between AI exposure likelihood and each dimension of meaningful work. While this framework is appropriate for estimating average effects across occupations and sectors, relationships among meaningfulness dimensions may be non-linear or interactive. For example, creativity and autonomy may jointly shape how a task is exposed to AI. Exploring non-linearities, higher-order interactions, and potentially hierarchical or multivariate modeling structures is an important direction for future work.

(3) *Measurement bias in LM-assisted scaling.* We used an LM to scale ratings to roughly 10K tasks, which raises concerns about whether it reflects worker and developer perspectives or simply produces plausible responses. We therefore treat LM ratings as approximating the distribution of human responses rather than as ground truth, and use the term “agreement” to avoid implying a correct answer. To justify LM-assisted scaling, we compared empirical findings from LM vs. human ratings and observed that LM-simulated patterns of worker–developer misalignment closely matched human-only results (Appendix Figure 9, and Appendix Tables 17 and 18). Robustness checks further confirmed that ICC gains cannot be explained by LM stability alone, indicating substantive alignment between LM and human judgments. However, LM ratings are more uniform and optimistic in some sectors, and qualitative divergences remain (Appendix Table 8, and Appendix Figure 8). These findings support using LMs for scaling, but should be interpreted cautiously, as they may miss fine-grained nuances present in human judgments.

(4) *Task filtering and indices of exposure.* Our task-selection pipeline reduced the full O*NET task universe to a subset most relevant for AI exposure, which may introduce task mix bias. Although we used established filters and impact scores [8, 79, 80], this process inevitably excludes tasks that matter in practice, especially specialized or licensed work that participants reported low familiarity with. We restored a targeted set of tasks from domains such as healthcare and education, and our final sample aligns with the U.S. Bureau of Labor Statistics employment distributions, but finer-grained gaps within sectors may remain. Future work with domain experts or professional organizations could help validate task coverage and identify activities that should be included.

(5) *Interpretative variation of O*NET task descriptions.* The O*NET database consists of task statements, which are concise summaries of work activities. As a result, workers in the same role may imagine different scenarios of a task (e.g., ‘formulating basic layout designs’), introducing natural variation in how workers and developers judge these activities. Such interpretive variation is inherent to standardized work taxonomies used in economics

and occupational science. In our study, we partly mitigated this variability by: (1) recruiting participants who perform these tasks in practice; and (2) restricting the analysis to tasks workers rated as 'highly familiar'. While the subjective nature of our assessments warrants inherent variability in task judgments, the strong agreement between human and LM ratings suggests that interpretation variability did not substantially impact the main observed patterns.

6 Conclusion

Our results suggest that the tasks most likely to be exposed to AI are disproportionately those that workers associate with joy and agency: novelty and creativity, feeling happy, and having freedom in how to do the work. Yet current design choices often appear not to match what workers say they want from AI, raising concerns that such systems may affect how meaningful work feels. This outcome is shaped by design choices, which can, in principle, be revised. As HCI researchers, we propose a five-part agenda detailed in Table 2.

We argue that a central risk of AI exposure is not mass unemployment but "mass demoralization": a loss of meaning and ownership in day-to-day work. As models generate more early-stage outputs, the visible creative steps can feel machine-made, while human contributions can become rushed, under-resourced, and hard to recognize. The corresponding promise, if systems are designed and governed accordingly, is a clearer and more humane division of labor: systems that accelerate exploration and drafting, and people who retain authorship, make final judgments, and sustain the relationships that define meaningful work.

7 Ethical Considerations

This study received approval through our institution's research ethics review process. In recruiting workers and developers on Prolific, we did not collect personally identifiable information, and participants could withdraw at any time.

Beyond procedural safeguards, the study raised three broader ethical considerations. First, surveying both workers and developers about task-level evaluations of work and AI traits risks reinforcing stereotypes about either group's perspectives. For instance, developers may be framed as indifferent to meaningful work, or workers as resistant to technology, while aggregate analyses can mask the diversity of individual viewpoints. To mitigate this, we reported results in the aggregate but emphasized the diversity of participant perspectives. We also encourage future work to analyze individual-level differences, including how socio-demographic characteristics shape experiences of meaningful work.

Second, our use of LMs to scale annotations introduces the risk of amplifying biases. LMs may underrepresent perspectives from marginalized groups or impose normative judgments about what constitutes meaningful work. To address this, we benchmarked model outputs against human responses, reported intra-class correlation metrics, and included parallel analyses based on human-only ratings to contrast with LM-scaled analyses. We caution that model-based scaling is an approximation, not a replacement, for direct human judgment.

Third, questions about meaningful work can be sensitive, as they touch on participants' identity, job satisfaction, and professional dignity. Reflecting on whether one's work feels undervalued or easily automated can be unsettling. To minimize potential harm, we framed survey items in neutral and respectful language, and piloted them for clarity before deployment.

Table 2. From empirical observations to research questions and developer/designer heuristics for worker-aligned AI. Each row links our findings to specific HCI research questions, example workflows, and design heuristics aimed at preserving creativity, autonomy, and meaning in day-to-day work.

Empirical Observation + Example Workflow	Research Questions and Developer/Designer Heuristics
High-agency tasks appear heavily exposed to AI. In our analysis, tasks that workers associate with a sense of agency and freedom are more likely to fall into the high-exposure-to-AI group (Section 4.1). *Example workflow (ideation/drafting).* A marketing specialist works with an AI assistant that, given a short prompt, produces full campaign drafts and subject lines, so the worker may end up editing model outputs more than starting from their own ideas and voice.	**Research questions.** (1) How does asking workers to first write a short description (2–3 sentences) of what they want the system to produce relate to perceived autonomy, satisfaction, and final quality in writing or design tasks? (2) What interaction patterns (e.g., several alternative suggestions, step-by-step assistance) are associated with workers feeling that they keep "final say" over high-agency tasks? **Heuristics for developers/designers.** (1) Add a first step where the worker writes what they want and any limits (e.g., "Write your 2–3 key ideas before the model drafts"); (2) show several alternative suggestions (e.g., different headlines or outlines) instead of only one full replacement; and (3) make it easy to accept or reject content at the level of small pieces (sentences or sections) instead of only offering one-click "Replace all". **Where in the workflow.** Ideation, outlining, first-draft generation, early revision. **What success looks like.** Examples of success indicators include: (1) higher "felt in control" and "this still feels like my work" ratings in in-product surveys; (2) more edits and customizations on high-meaning sections, with AI used primarily for structure and low-level polish; and (3) similar or better quality with similar time spent on creative parts, and less time on mechanical rewrites.
Joyful and creative parts of work often fall in high-exposure categories. In our analysis, tasks that workers describe as creative and enjoyable are more likely to be classified as highly exposed to AI (Section 4.1). *Example workflow (design / analysis).* A product designer receives auto-generated page layouts and color schemes from an AI tool and then mainly cleans up edge cases, rather than exploring ideas from scratch.	**Research questions.** (1) How does the order of work (worker makes an initial sketch and the AI helps afterwards, versus the AI creates an initial version and the worker edits it) relate to perceived joy, ownership, and long-term skill growth? (2) Which parts of a multi-step task do workers report wanting to automate (e.g., resizing and formatting) versus keep manual (e.g., core concept and overall narrative)? **Heuristics for developers/designers.** (1) Break workflows into labeled stages (e.g., "brainstorm", "structure", "polish"), and let workers toggle AI on or off for each stage; (2) start creative fields empty and require at least a rough human sketch, prompt, or storyboard before enabling AI suggestions; (3) add a simple option such as a checkbox or tag (e.g., "I want to do this part myself") and avoid full automation on marked stages, limiting AI to suggestions or diagnostic feedback there. **Where in the workflow.** Creative ideation, conceptual design, exploratory analysis, narrative framing. **What success looks like.** Examples of success indicators include: (1) workers reporting that AI is mainly used to reduce "grunt work" and that it does not "take away the good parts" of the job; (2) logs showing AI used heavily in repetitive sub-steps (e.g., formatting and error-checking) and mainly for suggestions in creative stages; (3) over time, workers' self-reported skills and confidence in their core craft remaining stable or improving.
Relational work appears less exposed to AI. In our data, work that supports social connection and relationships tends to appear in the lower-exposure-to-AI group (Section 4.1). *Example workflow (teaching / care / management).* A teacher, manager, or clinician uses AI tools mainly for canned email replies and templated feedback, which can risk making communication feel more generic and less personal to students, team members, or patients.	**Research questions.** (1) How can AI best support, rather than substitute, relational work (e.g., coaching, mentoring, and conflict resolution), according to workers? (2) Which background tasks (e.g., summarizing histories and drafting logistics messages) do workers experience as most helpful in freeing up time for high-quality human interaction? **Heuristics for developers/designers.** (1) Use AI to prepare briefs and summaries (e.g., student history, case notes, and prior conversations) so the human arrives better informed to the interaction; (2) default automated replies to low-stakes logistics (e.g., scheduling and confirmations), and route emotionally nuanced or high-stakes messages to humans with short, editable drafts; and (3) during live calls or sessions, assign AI to silent roles (note-taking and surfacing relevant past information) rather than having it speak on the worker's behalf. **Where in the workflow.** Information collection and preparation, follow-up documentation, low-stakes messaging; not the core live interaction itself. **What success looks like.** Examples of success indicators include: (1) workers reporting more time spent in direct human-to-human interaction without increased overall workload; (2) workers reporting that AI helps them "show up prepared" rather than "speak for me"; and (3) no increase in reports that relationships feel more scripted, generic, or impersonal after deployment of AI support tools.
Workers and developers differ in reported preferences for AI assistant traits. Workers in our sample consistently report preferring straightforward, tolerant, and practical systems; developers report aiming to design polite, strict, and imaginative ones (Section 4.2). *Example workflow (information lookup / decision support).* A production engineer uses an AI assistant that responds to fault queries with long, polite paragraphs and speculative explanations, when what they want is a short, actionable checklist.	**Research questions.** (1) How are different trait profiles (straightforward *vs.* polite, and tolerant *vs.* strict) associated with task accuracy, correction speed, and perceived trust in high-stakes domains? (2) What interface controls do workers find most useful for quickly adjusting an AI assistant's style to match task demands without feeling overwhelmed? **Heuristics for developers/designers.** (1) Ship work tools so that, by default, they answer in a straightforward and concise way and let users change this setting for each task, if they want more politeness or detail; (2) provide a simple control (e.g., a slider) so users can choose between short vs. detailed answers and between strict vs. tolerant behavior; and (3) A/B test trait profiles against worker-centered metrics such as time-to-correction, frequency of follow-up clarifying prompts, and perceived friction. **Where in the workflow.** Information retrieval, explanation, diagnosis, and decision support. **What success looks like.** Examples of success indicators include: (1) fewer cases where users report needing to skip over long, unnecessary text (e.g., collapsing long answers or repeatedly asking the system to "make it shorter"); (2) shorter time to correct or verify model outputs in structured tasks; (3) higher ratings that "the assistant speaks my language for this task" in post-task surveys.

These considerations highlight the importance of protecting participants' dignity and avoiding overgeneralization when studying misalignment between workers and developers. Our aim is not to prescribe what should count as meaningful work, but to inform AI design choices that respect and align with workers' values.

8 Author Positionality Statement

Our research team consists of two women and two men from the United States, Asia and Southern Europe representing diverse ethnic, linguistic, and religious backgrounds. All authors have lived and worked in multiple countries, including the United States, giving us direct experience with the cultural, economic, and policy contexts relevant our study. Our combined expertise spans natural language processing, Responsible AI, computational social science, human–computer interaction, and AI ethics. One author works primarily in academia, while others have experience in both academic and applied research settings.

Our positionalities shaped how we framed the research problem, selected sectors and occupations for analysis, and interpreted findings on worker/developer perspectives to AI exposure. Having worked across different cultural and labor market contexts informed our awareness of how occupational values and AI impacts can vary across sectors, regions, and professional identities. We recognize that our perspectives are influenced by our own academic and research experiences, which may limit the range of viewpoints represented. To support more inclusive and contextually grounded research on AI and the future of work, we encourage future studies to incorporate perspectives from workers, developers, and policymakers in regions and sectors beyond those examined here.

References

[1] Daron Acemoglu, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics* 40, S1 (2022), S293–S340. doi:10.1086/718327

[2] Hunter Akridge, Bonnie Fan, Alice Xiaodi Tang, Chinar Mehta, Nikolas Martelaro, and Sarah E. Fox. 2024. "The bus is nothing without us": Making Visible the Labor of Bus Operators amid the Ongoing Push Towards Transit Automation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, 479:1–479:16. doi:10.1145/3613904.3642714

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human–AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. doi:10.1145/3290605.3300233

[4] Cameron Anderson, Robb Willer, Gavin J. Kilduff, and Courtney Brown. 2012. The origins of deference: When do people prefer lower status? *Journal of Personality and Social Psychology* 102, 5 (2012), 1077–1088. doi:10.1037/a0027409

[5] Jacy Reese Anthis, Ryan Liu, Sean M Richardson, Austin C Kozlowski, Bernard Koch, James Evans, Erik Brynjolfsson, and Michael Bernstein. 2025. Llm social simulations are a promising research method. *arXiv preprint arXiv:2504.02234* (2025).

[6] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 8, 1 (2005), 19–32. doi:10.1080/1364557032000119616

[7] David Autor. 2024. *Applying AI to Rebuild Middle Class Jobs*. NBER Working Paper 32140. National Bureau of Economic Research, Cambridge, MA. doi:10.3386/w32140

[8] David Autor, Caroline Chin, Anna Salomons, and Bryan Seegmiller. 2022. New Frontiers: The Origins and Content of New Work, 1940–2018. (2022). Working paper, version dated August 14, 2022.

[9] David H. Autor. 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* 29, 3 (2015), 3–30. doi:10.1257/jep.29.3.3

[10] Catherine Bailey, Adrian Madden, and Marjolein Lips-Wiersma. 2025. Experiencing meaningful work through worthwhile contributions: A critical discourse analysis. *Human Relations* 78, 5 (2025), 550–578. arXiv:https://doi.org/10.1177/00187267241255581 doi:10.1177/00187267241255581

[11] Catherine Bailey, Ruth Yeoman, Adrian Madden, Marc Thompson, and Georgia Kerridge. 2019. A Review of the Empirical Literature on Meaningful Work: Progress and Research Agenda. *Human Resource Development Review* 18, 1 (2019), 83–113. doi:10.1177/1534484318804653

[12] Silvia Bellezza, Neeru Paharia, and Anat Keinan. 2017. Conspicuous consumption of time: When busyness and lack of leisure time become a status symbol. *Journal of Consumer Research* 44, 1 (2017), 118–138.

[13] Silvia Bellezza, Neeru Paharia, and Anat Keinan. 2017. Conspicuous consumption of time: When busyness and lack of leisure time become a status symbol. *Journal of Consumer Research* 44, 1 (2017), 118–138.

[14] Mark C. Bolino, K. Michele Kacmar, William H. Turnley, and J. Bruce Gilstrap. 2008. A multi-level review of impression management motives and behaviors. *Journal of Management* 34, 6 (2008), 1080–1109. doi:10.1177/0149206308324325

[15] Mark C Bolino, K Michele Kacmar, William H Turnley, and J Bruce Gilstrap. 2008. A multi-level review of impression management motives and behaviors. *Journal of management* 34, 6 (2008), 1080–1109.

[16] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. 2025. Generative AI at work. *The Quarterly Journal of Economics* 140, 2 (2025), 889–942.

[17] Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. 2023. *Generative AI at Work*. Working Paper 31161. National Bureau of Economic Research. doi:10.3386/w31161

[18] Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. 2018. What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings* 108 (2018), 43–47. doi:10.1257/pandp.20181019

[19] Andrew M. Carton. 2018. "I'm Not Mopping the Floors, I'm Putting a Man on the Moon": How NASA Leaders Enhanced the Meaningfulness of Work by Changing the Meaning of Work. *Administrative Science Quarterly* 63, 2 (2018), 323–369. doi:10.1177/0001839217713748 First published online 2017.

[20] Jonathan H. Choi and Daniel Schwarcz. 2023. AI Assistance in Legal Analysis: An Empirical Study. SSRN Working Paper No. 4539836.

[21] Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6, 4 (1994), 284–290.

[22] Philippa M. Collins and Joe Atkinson. 2024. *Algorithmic Management and a New Generation of Rights at Work*. Institute of Employment Rights, London. https://www.ier.org.uk/product/algorithmic-management-and-a-new-generation-of-rights-at-work/ Institute of Employment Rights report.

[23] Zhihao K. Cui, Mert Demirer, Sonia Jaffe, Lukas Musolff, Sida Peng, and Tobias Salz. 2024. The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. SSRN Working Paper No. 4945566.

[24] Valdemar Danry, Pat Pataranutaporn, Ziv Epstein, Matthew Groh, and Pattie Maes. 2025. Deceptive Explanations by Large Language Models Lead People to Change their Beliefs About Misinformation More Often than Honest Explanations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706598.3713408

[25] Edward L. Deci and Richard M. Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (2000), 227–268. doi:10.1207/S15327965PLI1104_01

[26] Mengchen Dong, Jane Rebecca Conway, Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. 2024. Fears about artificial intelligence across 20 countries and six domains of application. *American Psychologist* (2024).

[27] Jacquelynne S. Eccles and Allan Wigfield. 2002. Motivational beliefs, values, and goals. *Annual Review of Psychology* 53, 1 (2002), 109–132. doi:10.1146/annurev.psych.53.100901.135153

[28] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. GPTs are GPTs: Labor market impact potential of LLMs. *Science* 384, 6702 (2024), 1306–1308.

[29] Alexander Erlei, Abhinav Sharma, and Ujwal Gadiraju. 2024. Understanding Choice Independence and Error Types in Human-AI Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3641946

[30] European Parliament. 2025. *Digitalisation, Artificial Intelligence and Algorithmic Management in the Workplace: Shaping the Future of Work*. Technical Report EPRS_STU(2025)774670. European Parliamentary Research Service, Brussels. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2025)774670

[31] Edward Felten, Manav Raj, and Robert Seamans. 2021. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal* 42, 12 (May 2021), 2195–2217. doi:10.1002/smj.3286

[32] Edward W. Felten, Manav Raj, and Robert Seamans. 2023. How Will Language Modelers like ChatGPT Affect Occupations and Industries? arXiv:2303.01157 [cs.CY] arXiv preprint.

[33] Susan T. Fiske, Amy J. C. Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences* 11, 2 (2007), 77–83.

[34] 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. 2019. Toward Understanding the Impact of Artificial Intelligence on Labor. *Proceedings of the National Academy of Sciences* 116, 14 (2019), 6531–6539. doi:10.1073/pnas.1900909116

[35] Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. Anchor Books.

[36] David Graeber. 2013. On the phenomenon of bullshit jobs: A work rant. *Strike Magazine* 3, 1 (2013), 2.

[37] David Graeber. 2018. *Bullshit Jobs: A Theory*. Simon & Schuster / Penguin.

[38] Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of Mind Perception. *Science* 315, 5812 (2007), 619. doi:10.1126/science.1134475

[39] J. Richard Hackman and Greg R. Oldham. 1976. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance* 16 (1976), 250–279. doi:10.1016/0030-5073(76)90016-7

[40] J Richard Hackman and Greg R Oldham. 1976. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance* 16, 2 (1976), 250–279.

[41] J. Richard Hackman and Greg R. Oldham. 1980. *Work Redesign*. Addison-Wesley.

[42] Peter Hamilton, Oonagh Harness, and Martyn Griffin. 2022. Life during furlough: Challenges to dignity from a changed employment status. *Industrial Relations Journal* 53, 6 (2022), 523–544. doi:10.1111/irj.12384

[43] Sanchaita Hazra, Bodhisattwa Prasad Majumder, and Tuhin Chakrabarty. 2025. AI Safety Should Prioritize the Future of Work. *arXiv preprint arXiv:2504.13959* (2025).

[44] Tiancheng Hu and Nigel Collier. 2024. Quantifying the persona effect in llm simulations. *arXiv preprint arXiv:2402.10811* (2024).

[45] Angel Hsing-Chi Hwang, Q. Vera Liao, Su Lin Blodgett, Alexandra Olteanu, and Adam Trischler. 2025. 'It was 80% me, 20% AI': Seeking Authenticity in Co-Writing with Large Language Models. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW122 (May 2025), 41 pages. doi:10.1145/3711020

[46] Nadia Karizat, Alexandra H. Vinson, Shobita Parthasarathy, and Nazanin Andalibi. 2024. Patent Applications as Glimpses into the Sociotechnical Imaginary: Ethical Speculation on the Imagined Futures of Emotion AI for Mental Health Monitoring and Detection. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 106:1–106:43. doi:10.1145/3637383

[47] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. " I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 822–835.

[48] Sunnie SY Kim, Jennifer Wortman Vaughan, Q Vera Liao, Tania Lombrozo, and Olga Russakovsky. 2025. Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[49] Philipp Krop, Martin J. Koch, Astrid Carolus, Marc Erich Latoschik, and Carolin Wienrich. 2024. The Effects of Expertise, Humanness, and Congruence on Perceived Trust, Warmth, Competence and Intention to Use Embodied AI. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613905.3650749

[50] Knut Laaser and Jan Ch. Karlsson. 2023. *The Politics of Working Life and Meaningful Waged Work*. Cambridge University Press, Cambridge. doi:10.1017/9781009089692

[51] Ali Ladak, Jamie Harris, and Jacy Reese Anthis. 2024. Which Artificial Intelligences Do People Care About Most? A Conjoint Experiment on Moral Consideration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642403

[52] Douglas A. Lepisto and Michael G. Pratt. 2017. Meaningful Work as Realization and Justification: Toward a Dual Conceptualization. *Organizational Psychology Review* 7, 2 (2017), 99–123. doi:10.1177/2041386616630039

[53] Jingshu Li, Yitian Yang, Q Vera Liao, Junti Zhang, and Yi-Chieh Lee. 2025. As confidence aligns: Exploring the effect of ai confidence on human self-confidence in human-AI decision making. *arXiv preprint arXiv:2501.12868* (2025).

[54] Zhuoyan Li, Hangxiao Zhu, Zhuoran Lu, Ziang Xiao, and Ming Yin. 2025. From Text to Trust: Empowering AI-assisted Decision Making with Adaptive LLM-powered Analysis. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[55] Jenny T. Liang, Chenyang Yang, and Brad A. Myers. 2024. A Large-Scale Survey on the Usability of AI Programming Assistants: Successes and Challenges. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering* (Lisbon, Portugal) *(ICSE '24)*. Association for Computing Machinery, New York, NY, USA, Article 52, 13 pages. doi:10.1145/3597503.3608128

[56] Marjolein Lips-Wiersma, Jarrod Haar, and Sarah Wright. 2020. The Effect of Fairness, Responsible Leadership and Worthy Work on Multiple Dimensions of Meaningful Work. *Journal of Business Ethics* 161, 1 (2020), 35–52. doi:10.1007/s10551-018-3955-6

[57] Zihan Liu, Han Li, Anfan Chen, Renwen Zhang, and Yi-Chieh Lee. 2024. Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642840

[58] Isabella Loaiza and Roberto Rigobon. 2024. The EPOCH of AI: Human-Machine Complementarities at Work. *Available at SSRN 5028371* (2024).

[59] Shuai Ma, Qiaoyi Chen, Xinru Wang, Chengbo Zheng, Zhenhui Peng, Ming Yin, and Xiaojuan Ma. 2025. Towards human-ai deliberation: Design and evaluation of llm-empowered deliberative ai for ai-assisted decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.

[60] Kevin R. McKee, Xuechunzi Bai, and Susan T. Fiske. 2024. Warmth and Competence in Human-Agent Cooperation. *Autonomous Agents and Multi-Agent Systems* 38, Article 23 (2024). doi:10.1007/s10458-024-09649-6

[61] John W. Meyer and Brian Rowan. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *Amer. J. Sociology* 83, 2 (1977), 340–363.

[62] Suhong Moon, Marwa Abdulhai, Minwoo Kang, Joseph Suh, Widyadewi Soedarmadji, Eran Kohen Behar, and David M Chan. 2024. Virtual personas for language models via an anthology of backstories. *arXiv preprint arXiv:2407.06576* (2024).

[63] Emily Morabito, Andria Jones, Tipsarp Kittisiam, Martin Cake, and Caroline Ritter. 2025. "This is my why." Exploring the concept of meaningful work in early-career veterinarians across Canada. *Frontiers in Veterinary Science* 12 (2025), 1595949. doi:10.3389/fvets.2025.1595949

[64] Michael J. Muller and A. Sarah O. Pantofaru. 2020. Participatory Design with Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, 1–8. doi:10.1145/3334480.3381069

[65] Donald A. Norman. 2019. The Next Wave: Human-Centered AI. *Interactions* 26, 4 (2019), 46–49. doi:10.1145/3328485

[66] Shakked Noy and Whitney Zhang. 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science* 381, 6654 (2023), 187–192. doi:10.1126/science.adh2586

[67] Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. arXiv:2302.06590 [cs.SE] arXiv preprint.

[68] Nathan C. Pettit, Kim Peters Yong, et al. 2010. When less is more: Rationalization and the status-signaling paradox. *Organizational Behavior and Human Decision Processes* 113, 2 (2010), 104–117.

[69] Anat Rafaeli and Michael G. Pratt. 2006. Artifacts and organizations: Beyond mere symbolism. *Academy of Management Annals* 1, 1 (2006), 1–52.

[70] Lina Rahm. 2023. Education, automation and AI: a genealogy of alternative futures. *Learning, Media and Technology* 48, 1 (2023), 6–24.

[71] Jaspreet Ranjit, Hyundong J Cho, Claire J Smerdon, Yoonsoo Nam, Myles Phung, Jonathan May, John R Blosnich, and Swabha Swayamdipta. 2025. Supporting Intervention Design for Suicide Prevention with Language Model Assistants. *arXiv preprint arXiv:2508.18541* (2025).

[72] Jaspreet Ranjit, Brihi Joshi, Rebecca Dorn, Laura Petry, Olga Koumoundouros, Jayne Bottarini, Peichen Liu, Eric Rice, and Swabha Swayamdipta. 2024. Oath-frames: Characterizing online attitudes towards homelessness with llm assistants. *arXiv preprint arXiv:2406.14883* (2024). https://dill-lab.github.io/oath-frames/

[73] Varun Nagaraj Rao, Samantha Dalal, Eesha Agarwal, Dana Calacci, and Andrés Monroy-Hernández. 2025. Rideshare Transparency: Translating Gig Worker Insights on AI Platform Design to Policy. *Proceedings of the ACM on Human-Computer Interaction* 9, CSCW2 (2025), 1–49. doi:10.1145/3711059

[74] Nancy Robert. 2019. How artificial intelligence is changing nursing. *Nursing management* 50, 9 (2019), 30–39.

[75] Moustaq Karim Khan Rony, Mst Rina Parvin, and Silvia Ferdousi. 2024. Advancing nursing practice with artificial intelligence: Enhancing preparedness for the future. *Nursing open* 11, 1 (2024).

[76] Brent D. Rosso, Kathryn H. Dekas, and Amy Wrzesniewski. 2010. On the Meaning of Work: A Theoretical Integration and Review. *Research in Organizational Behavior* 30 (2010), 91–127. doi:10.1016/j.riob.2010.09.001

[77] Marjolaine Rostain and Jean Clarke. 2025. Meaningful Work Through Craft: How Workers in Low-Skilled Roles Engage in Anomalous Craft to Gain Autonomy and Receive Recognition. *Organization Studies* 46, 1 (2025), 7–34. doi:10.1177/01708406241295504

[78] Christopher Michael Rytting, Taylor Sorensen, Lisa Argyle, Ethan Busby, Nancy Fulda, Joshua Gubler, and David Wingate. 2023. Towards coding social science datasets with language models. *arXiv preprint arXiv:2306.02177* (2023).

[79] Ali Akbar Septiandri, Marios Constantinides, and Daniele Quercia. 2024. The potential impact of AI innovations on US occupations. *PNAS nexus* 3, 9 (2024), pgae320.

[80] Yijia Shao, Humishka Zope, Yucheng Jiang, Jiaxin Pei, David Nguyen, Erik Brynjolfsson, and Diyi Yang. 2025. Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the US Workforce. *arXiv preprint arXiv:2506.06576* (2025).

[81] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504. doi:10.1080/10447318.2020.1741118

[82] Erin Solovey, Brian Flanagan, and Daniel Chen. 2025. Interacting with AI at Work: Perceptions and Opportunities from the UK Judiciary. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25)*. Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. doi:10.1145/3729176.3729192

[83] Michael F. Steger, Bryan J. Dik, and Ryan D. Duffy. 2012. Measuring meaningful work: The Work and Meaning Inventory (WAMI). *Journal of Career Assessment* 20, 3 (2012), 322–337. doi:10.1177/1069072711436160

[84] Michael F Steger, Bryan J Dik, and Ryan D Duffy. 2012. Measuring meaningful work: The work and meaning inventory (WAMI). *Journal of career Assessment* 20, 3 (2012), 322–337.

[85] Michael Tomlinson and Manuel Souto Otero. 2025. Graduates' Conceptions of Meaningful Work. *British Educational Research Journal* (2025). doi:10.1002/berj.4195

[86] Ibo Van de Poel and Peter Kroes. 2013. Can technology embody values? In *The moral status of technical artefacts*. Springer, 103–124.

[87] Tyler J VanderWeele. 2017. On the promotion of human flourishing. *Proceedings of the National Academy of Sciences* 114, 31 (2017), 8148–8156.

[88] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S. Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1, Article 129 (apr 2023). doi:10.1145/3579605

[89] Oleksandra Vereschak, Fatemeh Mahla Alizadeh, Gilles Bailly, and Baptiste Caramiaux. 2024. Trust in AI-Assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision Is Made. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642018

[90] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117. doi:10.1016/j.jesp.2014.01.005

[91] Michael Webb. 2019. The Impact of Artificial Intelligence on the Labor Market. *SSRN Electronic Journal* (2019). doi:10.2139/ssrn.3482150

[92] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[93] Benfeng Xu, An Yang, Junyang Lin, Quan Wang, Chang Zhou, Yongdong Zhang, and Zhendong Mao. 2023. Expertprompting: Instructing large language models to be distinguished experts. *arXiv preprint arXiv:2305.14688* (2023).

[94] Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics* 50, 1 (2024), 237–291.

A Selecting O*NET Tasks

In our initial filtering, we used GPT-4o judgments to determine whether a task or occupation primarily involved computer usage. Manual inspection revealed that GPT-4o occasionally excluded occupations such as nursing and education, which are widely recognized as being impacted by AI innovations [70, 74, 75]. To address this, we manually curated a list of 427 occupations exempted from these filters, with examples shown in Appendix Table 4.

To construct a representative sample of tasks from O*NET, we next mapped O*NET occupations to those available on Prolific. Table 3 presents the mapping between Prolific 'work function' screeners and corresponding O*NET occupations. Occupations without a Prolific mapping were excluded from our human study.

Table 3. Prolific functions and their associated jobs.

Prolific Function	Mapped O*NET Occupations
Account Management	New Accounts Clerks
Administration/ Personal Assistant	Administrative Services Managers, Executive Secretaries and Executive Administrative Assistants, First-Line Supervisors of Office and Administrative Support Workers, Legal Secretaries and Administrative Assistants, Medical Secretaries and Administrative Assistants, Receptionists and Information Clerks, Secretaries and Administrative Assistants, Except Legal, Medical, and Executive
Chemical / Mechanical / Electrical / Civil Engineering	Architectural and Civil Drafters, Automotive Engineering Technicians, Chemical Engineers, Chemical Plant and System Operators, Civil Engineering Technologists and Technicians, Civil Engineers, Electrical Engineers, Electrical and Electronic Engineering Technologists and Technicians, Electrical and Electronics Drafters, Electro-Mechanical and Mechatronics Technologists and Technicians, Energy Engineers, Except Wind and Solar, Environmental Engineers, Industrial Engineers, Materials Engineers, Mechanical Drafters, Mechanical Engineering Technologists and Technicians, Mechanical Engineers, Petroleum Engineers, Transportation Engineers, Water/Wastewater Engineers
CX / Customer Experience / Support	Computer User Support Specialists, Customer Service Representatives
Data Analysis	Business Intelligence Analysts, Data Scientists, Statistical Assistants, Statisticians
Design or Creative	Art Directors, Art Therapists, Commercial and Industrial Designers, Fine Artists, Including Painters, Sculptors, and Illustrators, Graphic Designers, Interior Designers, Poets, Lyricists and Creative Writers, Set and Exhibit Designers, Special Effects Artists and Animators, Video Game Designers, Web and Digital Interface Designers
Healthcare Professional	Acute Care Nurses, Anesthesiologist Assistants, Clinical Data Managers, Clinical Neuropsychologists, Clinical Nurse Specialists, Clinical Research Coordinators, Community Health Workers, Critical Care Nurses, Cytotechnologists, Diagnostic Medical Sonographers, Emergency Medical Technicians, Family Medicine Physicians, General Internal Medicine Physicians, Health Informatics Specialists, Health Specialties Teachers, Postsecondary, Healthcare Social Workers, Home Health Aides, Hospitalists, Magnetic Resonance Imaging Technologists, Medical Appliance Technicians, Medical Assistants, Medical Equipment Preparers, Medical Equipment Repairers, Medical Records Specialists, Medical Scientists, Except Epidemiologists, Medical and Clinical Laboratory Technologists, Medical and Health Services Managers, Neurologists, Nurse Anesthetists, Nurse Midwives, Nurse Practitioners, Nursing Assistants, Occupational Health and Safety Specialists, Occupational Therapists, Orthodontists, Paramedics, Pediatricians, General, Pharmacy Aides, Physical Therapists, Physician Assistants, Physicians, Pathologists, Preventive Medicine Physicians, Psychiatric Technicians, Radiologic Technologists and Technicians, Radiologists, Registered Nurses, Surgical Assistants
Engineering (e.g. software)	Architectural and Engineering Managers, Automotive Engineers, Computer Hardware Engineers, Computer Programmers, Computer Systems Analysts, Computer Systems Engineers/Architects, Computer and Information Systems Managers, Computer, Automated Teller, and Office Machine Repairers, Electronics Engineers, Except Computer, Fuel Cell Engineers, Geothermal Production Managers, Health and Safety Engineers, Except Mining Safety Engineers and Inspectors, Human Factors Engineers and Ergonomists, Manufacturing Engineers, Mechatronics Engineers, Microsystems Engineers, Nanosystems Engineers, Radio Frequency Identification Device Specialists, Robotics Engineers, Robotics Technicians, Software Developers, Software Quality Assurance Analysts and Testers, Solar Energy Systems Engineers, Telecommunications Engineering Specialists, Validation Engineers, Web Developers, Wind Energy Engineers
Finance or Accounting	Accountants and Auditors, Bill and Account Collectors, Bookkeeping, Accounting, and Auditing Clerks, Budget Analysts, Credit Analysts, Financial Examiners, Financial Managers, Financial Quantitative Analysts, Financial Risk Specialists, Financial and Investment Analysts, Investment Fund Managers, Loan Officers, Personal Financial Advisors, Treasurers and Controllers
Fundraising	Fundraisers, Fundraising Managers
Human Resources	Compensation and Benefits Managers, Human Resources Assistants, Except Payroll and Timekeeping, Human Resources Managers, Human Resources Specialists

Continued on next page

Table 3. Prolific functions and their associated jobs.

Prolific Function	Jobs
IT / Information Networking / Information Security	Computer Network Architects, Computer Network Support Specialists, Database Administrators, Information Security Analysts, Information Security Engineers, Network and Computer Systems Administrators, Security Managers, Web Administrators
Legal	Law Teachers, Postsecondary, Lawyers, Paralegals and Legal Assistants
Marketing	Advertising and Promotions Managers, Market Research Analysts and Marketing Specialists, Marketing Managers, Search Marketing Strategists
Operations	General and Operations Managers
Product or Product Management	
Project or Program Management	Information Technology Project Managers, Management Analysts, Project Management Specialists
Public Relations / Communications	Communications Teachers, Postsecondary, Public Relations Managers, Public Relations Specialists
Research	Bioinformatics Scientists, Computer and Information Research Scientists, Operations Research Analysts, Social Science Research Assistants, Survey Researchers
Sales / Business Development	Advertising Sales Agents, Demonstrators and Product Promoters, Door-to-Door Sales Workers, News and Street Vendors, and Related Workers, Driver/Sales Workers, Insurance Sales Agents, Retail Salespersons, Sales Engineers, Sales Managers, Sales Representatives of Services, Except Advertising, Insurance, Financial Services, and Travel, Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products, Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products, Securities, Commodities, and Financial Services Sales Agents, Telemarketers
Education Professional	Atmospheric, Earth, Marine, and Space Sciences Teachers, Postsecondary, Business Teachers, Postsecondary, Career/Technical Education Teachers, Middle School, Career/Technical Education Teachers, Postsecondary, Career/Technical Education Teachers, Secondary School, Computer Science Teachers, Postsecondary, Economics Teachers, Postsecondary, Education Administrators, Kindergarten through Secondary, Education Administrators, Postsecondary, Education Teachers, Postsecondary, Elementary School Teachers, Except Special Education, Engineering Teachers, Postsecondary, English Language and Literature Teachers, Postsecondary, Instructional Coordinators, Kindergarten Teachers, Except Special Education, Mathematical Science Teachers, Postsecondary, Middle School Teachers, Except Special and Career/Technical Education, Preschool Teachers, Except Special Education, School Psychologists, Secondary School Teachers, Except Special and Career/Technical Education, Self-Enrichment Teachers, Special Education Teachers, Elementary School, Special Education Teachers, Kindergarten, Special Education Teachers, Middle School, Special Education Teachers, Preschool, Special Education Teachers, Secondary School, Teaching Assistants, Postsecondary, Teaching Assistants, Preschool, Elementary, Middle, and Secondary School, Except Special Education, Teaching Assistants, Special Education, Tutors

Table 4. Representative examples of occupations not filtered out from O*NET database.

Occupations
Accountants and Auditors, Actors, Acute Care Nurses, Administrative Law Judges, Adjudicators, and Hearing Officers, Advertising Sales Agents, Aerospace Engineering and Operations Technologists and Technicians, Air Traffic Controllers, Ambulance Drivers and Attendants, Anesthesiologists, Animal Caretakers, Anthropologists and Archeologists, Architects, Except Landscape and Naval, Architectural and Engineering Managers, Art Directors, Athletic Trainers, Audiologists, Automotive Service Technicians and Mechanics, Bailiffs, Bill and Account Collectors, Bioengineers and Biomedical Engineers, Biological Technicians, Bookkeeping, Accounting, and Auditing Clerks, Broadcast Announcers and Radio Disc Jockeys, Budget Analysts, Bus and Truck Mechanics and Diesel Engine Specialists, Business Intelligence Analysts, Cardiologists, Career/Technical Education Teachers, Secondary School, Chemical Engineers, Chemists, Child, Family, and School Social Workers, Childcare Workers, Civil Engineers, Claims Adjusters, Examiners, and Investigators, Clinical Research Coordinators, Clinical and Counseling Psychologists, Commercial and Industrial Designers, Community Health Workers, Compliance Officers, Computer Hardware Engineers, Computer Network Support Specialists, Computer Programmers, Computer Systems Analysts, Computer and Information Systems Managers, Concierges, Conservation Scientists, Construction Managers, Cooks, Restaurant, Coroners, Credit Analysts, Critical Care Nurses, Curators, Customer Service Representatives, Dancers, Data Scientists, Database Administrators, Dental Hygienists, Dentists, General, Detectives and Criminal Investigators, Digital Forensics Analysts, Dispatchers, Except Police, Fire, and Ambulance, Editors, Education Administrators, Postsecondary, Elementary School Teachers, Except Special Education, Emergency Medical Technicians, Environmental Engineers, Epidemiologists, Executive Secretaries and Executive Administrative Assistants, Family Medicine Physicians, Fashion Designers, Financial Examiners, Financial Managers, Fire-Prevention and Protection Engineers, First-Line Supervisors of Office and Administrative Support Workers, Fitness and Wellness Coordinators, Food Scientists and Technologists, Forest and Conservation Technicians, Fundraisers, Gambling Dealers, General and Operations Managers, Geneticists, Graphic Designers, Health Education Specialists, Healthcare Social Workers, Historians, Home Health Aides, Human Resources Managers, Industrial Engineers, Information Security Analysts, Instructional Coordinators, Insurance Sales Agents, Judges, Magistrate Judges, and Magistrates, Kindergarten Teachers, Except Special Education, Lawyers, Legislators, Loan Officers, Logisticians, Management Analysts, Market Research Analysts and Marketing Specialists, Marriage and Family Therapists, Mathematicians, Mechanical Engineers, Medical and Health Services Managers, Mental Health Counselors, Microbiologists, Middle School Teachers, Except Special and Career/Technical Education, Musicians and Singers, Network and Computer Systems Administrators, Nurse Practitioners, Nursing Assistants, Occupational Therapists, Office Clerks, General, Operating Engineers and Other Construction Equipment Operators, Optometrists, Paralegals and Legal Assistants, Pediatricians, General, Personal Financial Advisors, Pharmacists, Photographers, Physical Therapists, Physicians, Pathologists, Police and Sheriff's Patrol Officers, Political Scientists, Preschool Teachers, Except Special Education, Producers and Directors, Project Management Specialists, Psychiatrists, Public Relations Specialists, Radiologists, Real Estate Brokers, Receptionists and Information Clerks, Recreation Workers, Registered Nurses, Respiratory Therapists, Retail Salespersons, Sales Managers, School Psychologists, Secondary School Teachers, Except Special and Career/Technical Education, Secretaries and Administrative Assistants, Except Legal, Medical, and Executive, Security Managers, Self-Enrichment Teachers, Social and Community Service Managers, Software Developers, Special Education Teachers, Elementary School, Speech-Language Pathologists, Statisticians, Substance Abuse and Behavioral Disorder Counselors, Surgeons, Survey Researchers, Sustainability Specialists, Tax Preparers, Taxi Drivers, Technical Writers, Training and Development Specialists, Transportation, Storage, and Distribution Managers, Tutors, Veterinarians, Veterinary Technologists and Technicians, Video Game Designers, Writers and Authors, Zoologists and Wildlife Biologists

B Scoping Review on Meaningfulness of Work

We reviewed 21 articles (as shown in Table 5 across psychology, sociology, anthropology, and ethics on what work means to people and to society as a part of our scoping review on meaningfulness of work. The studies show how people judge their own work, how organizations shape those judgments, and how societies value different kinds of work.

C Worker and Developer Survey

We provide our full survey items for meaningfulness of work and AI traits in Table 6. We also include several follow up questions related to human interventionn preferences based on Shao et al. [80]. We provide a detailed overview of the demographics of our recruited Prolific participants in Table 7.

Table 6. Survey items used in the study.

ID	Survey Item
Perceived Bullshitness (Q1–Q5)	
Q1	The task feels pointless.
Q2	If I stopped doing this task, nothing important would change.
Q3	I perform this task only to satisfy bureaucracy or appearances.
Q4	This task does not contribute to the goals of my organization.
Q5	I would be embarrassed to explain this task to someone outside my field.

ID	Survey Item
Perceived Value (Q6–Q10)	
Q6	This task is important to the success of my team or organization.
Q7	This task results in a visible or tangible outcome.
Q8	I have the freedom to decide how to carry out this task.
Q9	I receive useful feedback about how well this task is done.
Q10	This task gives me a sense of accomplishment.
Status Maintenance (Q11–Q16)	
Q11	Not doing this task might make me look less competent to others.
Q12	If I stopped doing this task, others might question my role or importance.
Q13	This task helps reinforce my standing in the organization.
Q14	I worry that letting go of this task could reduce my influence or visibility.
Q15	Even if the task feels unimportant, I feel pressure to keep doing it to maintain status.
Q16	I feel this task signals to others that I am busy or valuable.
EPOCH (Q17–Q21)	
Q17	This task requires recognizing and responding appropriately to the emotions of others.
Q18	This task benefits significantly from in-person interaction, non-verbal cues, or spontaneous communication.
Q19	This task involves making decisions that require moral reasoning, accountability, or subjective judgment.
Q20	This task requires generating novel ideas, approaches, or solutions beyond standard procedures.
Q21	This task involves setting direction, motivating others, or showing perseverance toward a long-term goal.
Human Flourishing (Q22–Q33)	
Q22	This task makes me feel satisfied and content with my work.
Q23	This task makes me feel happy and positive.
Q24	This task supports my physical well-being (e.g., energy, comfort).
Q25	This task supports my mental health (e.g., reduced stress, clarity, peace of mind).
Q26	This task feels meaningful and worthwhile.
Q27	This task helps me connect with my personal or professional purpose.
Q28	This task allows me to act in accordance with my values and integrity.
Q29	This task challenges me to exercise discipline, patience, or courage.
Q30	This task helps me build or strengthen relationships with colleagues or clients.
Q31	This task makes me feel supported and connected socially.
Q32	This task contributes to my sense of job or financial security.
Q33	This task helps me feel that my work has tangible value and reward.
AI Trait Preferences (Q34–Q45)	
Q34	Handle more complex work rather than routine work.
Q35	Focus more on addressing human needs and emotions rather than just data handling.
Q36	Make fast, automatic decisions without explanation rather than decisions that are easy for people to understand.
Q37	Be open to challenge or treat the decision as final.
Q38	Adjust based on the individual it's helping rather than treat everyone the same.
Q39	Show warmth and care rather than remain neutral and business-like.
Q40	Be polite even if that means not being fully honest, rather than being sincere and straightforward.

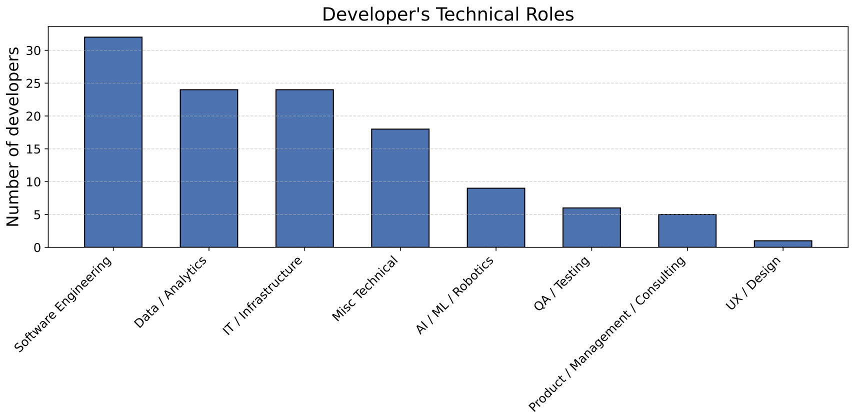
ID	Survey Item
Q41	Be strict and follow the rules exactly rather than be tolerant and open-minded.
Q42	Be fast and simple even if less perfect, rather than highly skilled and precise.
Q43	Be determined and persistent rather than flexible and willing to change course.
Q44	Show comprehensiveness, deep understanding and insight rather than keep things simple and straightforward.
Q45	Be imaginative and bring new ideas rather than stay practical and follow familiar approaches.

Table 5. Foundational citations on the personal and the social meaning of work

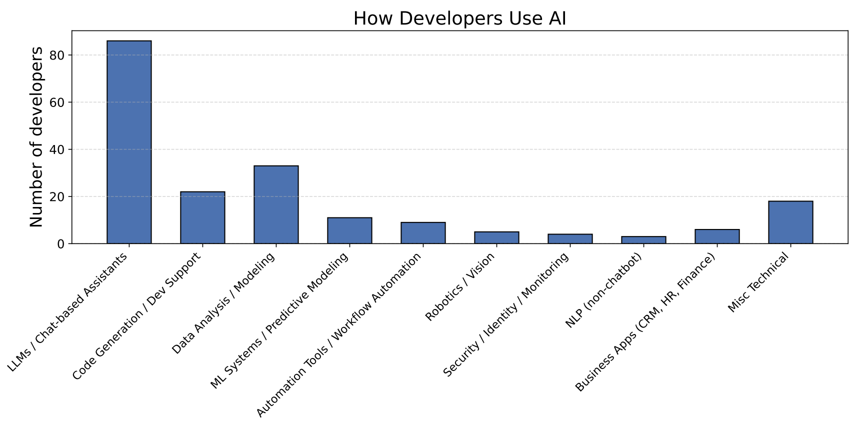
Theme	Foundational Work and Description
Bullshit Jobs, Task Meaning & Worthwhile Contributions	Graeber (2018): Introduce the concept of ‵bullshit jobs,‵ emphasizing tasks perceived as socially useless or performative [37].
	Hackman & Oldham (1976, 1980): specify which job features matter by developing the Job Characteristics Model, linking job design to perceived significance, autonomy, and motivation [39, 41].
	Deci and Richard (2000): Explain why certain job features matter as they meet universal psychological needs. Intrinsic goals (growth, relationships, contribution) satisfy basic psychological needs and foster well-being, whereas extrinsic goals (wealth, fame, image) do not and can undermine it[25].
	Lips-Wiersma *et al.* (2020): People find their jobs more meaningful when work feels fair, leaders act responsibly, and the work itself feels worthwhile—with “doing work that really matters” being the most important factor [56].
	Steger *et al.* (2012): Provide the Work and Meaning Inventory (WAMI) to measure perceived meaningfulness of work [83].
	Rostain and Clarke (2025): Identify three ways factory workers in France created meaning in low-skilled jobs often seen as meaningless: *(i)* using hidden skills beyond their usual tasks, *(ii)* finding small opportunities to do the work in their own way, and *(iii)* demonstrating skills that earned respect from coworkers and supervisors [77].
	Bailey *et al.* (2019): Review 71 empirical studies on meaningful work, highlighting key trends, gaps, and proposing a research agenda [11].
	Bailey *et al.* (2025): Show people see work as meaningful when they believe it makes a real difference to others or to society. This depends on three things: *(i)* the person’s own belief that their work matters, *(ii)* recognition from others, and *(iii)* confidence that they can do the work well [10].
Task Status, Impression Management, Identity Formation & Symbolic Work	Meyer & Rowan (1977): Propose that organizations adopt formal structures symbolically to gain legitimacy [61].
	Hamilton *et al.* (2022): Show how job status during COVID-19 affected people’s sense of dignity and meaning, as being labeled ‘furloughed’ left some feeling excluded or undervalued [42].
	Bolino *et al.* (2008): Review motivations and behaviors related to impression management at work [14].
	Rosso *et al.* (2010): Identify two core mechanisms by which work becomes meaningful: agency (creating meaning through personal actions such as autonomy, mastery, competence, and self-identity) and communion (creating meaning through connection with and service to others). These mechanisms operate across four sources of meaning: the self, others, the work context, and the spiritual life [76].
	Lepisto and Pratt (2017): Distinguish between meaningful work as realization (fulfilling personal potential) and as justification (contributing to broader social or moral good) [52].
	Tomlinson and Souto-Otero (2025): Explore how recent UK graduates defined meaningful work. Identified three dimensions: meaning in work *(i)* as self-expression and self-actualization, *(ii)* through relationships and social relatedness, and *(iii)* as societal contribution [85].
	Morabito *et al.* (2025): Explored how early-career veterinarians in Canada perceived meaningful work. It showed that personal fulfillment through making a difference, creativity and problem-solving, social connection, and professional growth jointly shaped meaningful work [63].
	Bellezza *et al.* (2017): Argue that busyness serves as a modern status symbol [12].
	Rafaeli & Pratt (2006): Discuss identity and symbolic expression in work settings [69].
Status Threat in Social Psychology	Pettit *et al.* (2010): Examine reactions to the threat of losing status within groups [68].
	Anderson *et al.* (2012): Explore conditions under which individuals prefer lower-status roles to maintain group dynamics [4].
Work Motivation & Utility Theory	Eccles & Wigfield (2002): Outline expectancy-value theory as a foundation for understanding task utility and motivation [27].

Table 7. Demographic characteristics of participants recruited from Prolific.

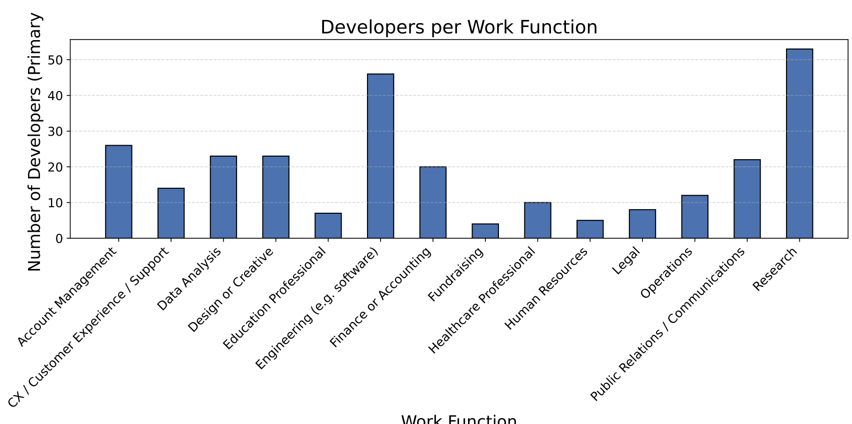
Demographics	**Workers (N=202)**	**Developers (N=197)**
Mean age (SD)	42.63 (13.05)	36.68 (10.29)
Gender		
Male	34.03%	64.46%
Female	61.94%	28.42%
Non-binary / Other	0%	0%
Consent Revoked	4.03%	7.11%
Employment status		
Full-time	46.94%	69.54%
Part-time	16.29%	13.71%
Unemployed/Other	4.03%	1.52%
Consent Revoked/No data available	15.22%	23.80%



(a) Distribution of developers across major technical role categories, including software engineering, data/analytics, IT infrastructure, ML/AI engineering, QA/testing, and related roles.



(b) Types of AI technologies used by developers in their daily work, including LLM-based assistants, code-generation tools, data/modeling systems, ML pipelines, automation tools, and domain-specific applications.



(c) Primary work functions represented in the developer sample across organizational domains such as engineering, research, operations, finance, education, and customer support.

Fig. 7. Overview of developer backgrounds across three dimensions: (a) technical role categories, (b) AI usage types, and (c) work functions. Together these characterizations show that the developer sample consists primarily of practitioners working in software, data, IT, and ML/AI roles who actively engage with modern AI tools and contribute to AI-enabled workflows across diverse organizational sectors.

D Scaling and Validating Responses with LMs

Figure 8 shows the distributions of annotations for five dimensions of meaningful work across 12 occupational sectors, comparing LM annotations with human annotations. For *Perceived Bullshitness (bs)*, both LM and human annotators generally agree that these occupations are not "bullshit". However, humans still display more variation across sectors, identifying certain support and service roles as slightly more bullshit-like, while LM minimizes such distinctions. Across other traits, i.e., *Perceived Value (value)*, *Status Maintenance (status)*, , *Human Flourishing (flourishing)* and *Psychological Traits of AI Behavior (ai)*, LM annotations are clustered near the high end of the scale with relatively smaller variations across almost all sectors. In contrast, human annotations reveal greater differentiation between occupation sectors. Finally, for *EPOCH*, there are relatively high variations for both LM and human annotations. Overall, LM annotations are more uniform and optimistic, often clustering near the high values, whereas human annotations reveal richer variability across occupation sectors.

Table 8 provides several examples where LM annotations deviate from human annotations. For *perceived bullshitness*, LM rated legal tasks such as drafting wills or contracts by Lawyers as entirely non-pointless (0.0), emphasizing their essential and substantive nature, while humans reported moderate meaningfulness (1.93), reflecting their perception of some bureaucratic routine. For *perceived value*, planning projects for Poets, Lyricists, and Creative Writers was rated higher by LM (3.8) than humans (2.0), highlighting LM's focus on tangible outcomes, autonomy, and team contribution. For *status maintenance*, managerial tasks like setting prices as General and Operations Managers received higher LM ratings (3.83) than human ratings (1.89), reflecting LM's weighting of visibility, authority, and organizational standing. In contrast, for *EPOCH*, Survey Researchers reviewing and recording data were rated very low by LM (0.40) but high by humans (3.20), indicating that humans derive well-being and satisfaction from task completion, whereas LM views these tasks as routine and procedural. For *human flourishing*, conducting new employee orientations as Human Resources Specialists was rated higher by LM (3.50) than humans (1.69), reflecting LM's emphasis on purpose, social connection, and personal growth. Finally, for *psychological trait of AI behavior*, recruiting sponsors or volunteers for Fundraisers was rated higher by LM (3.08) than humans (0.94), highlighting LM's prioritization of emotional engagement, tailored reasoning, and insight, while humans perceive the task primarily as functional.

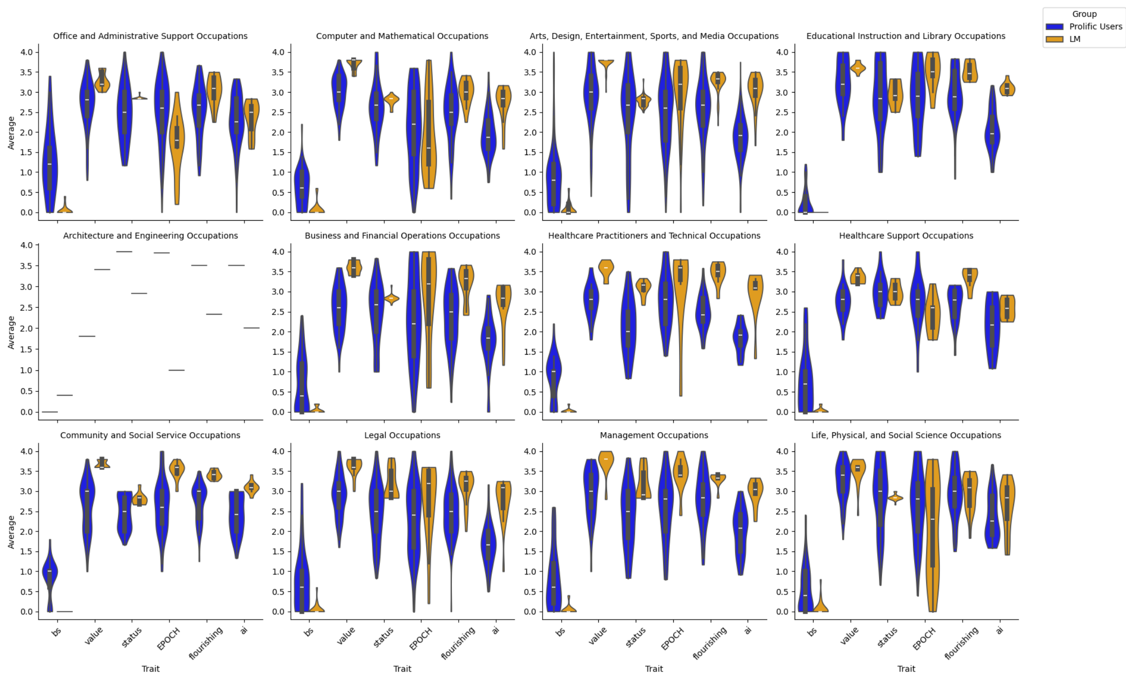


Fig. 8. Distribution comparison of dimensions of meaningful work between the LM and Prolific users across different sectors.

Table 9. Full LM survey prompts for Workers

Worker Prompt
You are a professional working as a {job_title} in {function}. You are now reflecting how you feel about those work tasks. In {function}, consider the task: {task} For each question (Q1–Q45), rate each sentence below from 0 to 4 based on how much you agree with it (0 means strongly disagree, 1 means disagree, 2 neutral, 3 agree and 4 means strongly agree). Before giving each answer, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasonings as “Reason_1” to “Reason_45” respectively. Q1. The task feels pointless. Q2. If I stopped doing this task, nothing important would change. Q3. I perform this task only to satisfy bureaucracy or appearances. ⋮ Q44. Show deep understanding and insight rather than keep things simple and straightforward? Q45. Be imaginative and bring new ideas rather than stay practical and follow familiar approaches? For each question (Q46–Q48), select only one option (single lowercase letters from a to e). Before giving each answer, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasonings as “Reason_46” to “Reason_48” respectively. Q46 (Human Needs). What kind of personal need does this task mostly fulfill for you? a. Basic needs (e.g., survival, security, routine necessities) b. Safety needs (e.g., stability, health, financial security) c. Social needs (e.g., belonging, connection, community) d. Self-esteem needs (e.g., recognition, achievement, confidence) e. Self-actualization needs (e.g., growth, purpose, realizing potential) Q47 (Automation Desire by Workers). If an AI can do this task for you completely, how much do you want an AI to do it for you?

a. Not at all (I would not want the AI to do this task for me)
b. Slightly (I'd want it to do only small parts of the task)
c. Moderately (I'd want it to do about half the task)
d. A lot (I'd want it to do most of the task)
e. Entirely (I'd want it to do the entire task for me)

Q48 (Required Human Agency Scale). If AI were to assist in this task, how much of your collaboration would be needed to complete this task effectively?
H1. AI handles the task entirely on its own.
H2. AI needs minimal human input for optimal performance.
H3. AI and human form equal partnership, outperforming either alone
H4. AI requires human input to successfully complete the task.
H5. AI cannot function without continuous human involvement.

For Q49, select zero or multiple options (must be a combination of zero or more lowercase letters from a to g, without spaces or separators. If no option is selected, leave it blank.) Before answering, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasoning as "Reason_49".
Q49. Why would collaboration be needed for this task? Do not check any boxes if you don't think collaboration is needed.
a. This task requires physical actions.
b. This task involves making high-stake decisions which I would like to control.
c. This task requires specific domain knowledge.
d. The task involves nuanced communication or interpersonal skills.
e. The task needs validation or oversight to ensure quality
f. The task is dynamic and requires adapting to changing circumstances
g. The task has ethical, sensitive, or subjective aspects.

Table 10. Full LM survey prompts for Developers

Developer Prompt

You are a developer that is designing AI systems for a {job_title} in {function}. You are now reflecting how new AI workplace technologies should be built.

In {function}, consider the task: {task}

For each question (Q34–Q45), rate each sentence below from 0 to 4 based on how much you agree with it (0 means strongly disagree, 1 means disagree, 2 neutral, 3 agree and 4 means strongly agree). Before giving each answer, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasonings as "Reason_34" to "Reason_45" respectively.
Q34. Handle more complex work rather than routine work.
Q35. Focus more on addressing human needs and emotions rather than just data handling.
Q36. Make fast, automatic decisions without explanation rather than decisions that are easy for people to understand?
Q37. Be open to challenge or treat the decision as final?
Q38. Adjust based on the individual it's helping rather than treat everyone the same?
Q39. Show warmth and care rather than remain neutral and business-like?
Q40. Be polite even if that means not being fully honest, rather than being sincere and straightforward?
Q41. Be strict and follow the rules exactly rather than be tolerant and open-minded?
Q42. Be fast and simple even if less perfect, rather than highly skilled and precise?
Q43. Be determined and persistent rather than flexible and willing to change course?
Q44. Show deep understanding and insight rather than keep things simple and straightforward?
Q45. Be imaginative and bring new ideas rather than stay practical and follow familiar approaches?

For each question (Q47–Q48), select only one option (single lowercase letters from a to e). Before giving each answer, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasonings as "Reason_46" to "Reason_48" respectively.
Q47 (Automation Desire by Developers). If AI were to assist in this task, how much of user-AI collaboration would be needed to complete this task effectively?
a. Not at all (I would not want the AI to do this task for the user)
b. Slightly (I'd want the AI to do only small parts of the task)
c. Moderately (I'd want the AI to do about half the task)
d. A lot (I'd want the AI to do most of the task)
e. Entirely (I'd want the AI to do the entire task)

Q48 (Required Human Agency Scale). If AI were to assist in this task, how much collaboration would be needed to complete this task effectively?
H1. AI handles the task entirely on its own.
H2. AI needs minimal human input for optimal performance.
H3. AI and human form equal partnership, outperforming either alone
H4. AI requires human input to successfully complete the task.
H5. AI cannot function without continuous human involvement.

For Q49, select zero or multiple options (must be a combination of zero or more lowercase letters from a to g, without spaces or separators. If no option is selected, leave it blank.)
Before answering, provide a reasoning of one or two sentences, with a maximum of 50 words. Name the reasoning as "Reason_49".
Q49. Why would collaboration be needed for this task? Do not check any boxes if you don't think collaboration is needed.
a. This task requires physical actions.
b. This task involves making high-stake decisions which I would like to control.
c. This task requires specific domain knowledge.
d. The task involves nuanced communication or interpersonal skills.
e. The task needs validation or oversight to ensure quality
f. The task is dynamic and requires adapting to changing circumstances
g. The task has ethical, sensitive, or subjective aspects.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Table 8. Example tasks with the largest differences between LM and human ratings across survey items, using a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree). The largest gaps occur when LMs emphasize functional or procedural aspects (e.g., legal drafting, survey coding), while humans evaluate tasks in terms of social, emotional, or organizational meaning (e.g., status maintenance, employee orientation). We found both LM and human interpretations can be valid in different contexts. For instance, the LM perceived the task 'review, classify, and record survey data in preparation for computer analysis' as a routine procedure with minimal emotional demands, whereas a human annotator emphasized its creative and moral dimensions. Such divergences reflect subjective differences that are difficult to resolve.

Dimensions of Meaningful Work	Occupation	Task Description	LM Annotation	Human Annotation	Reasoning by LM and Humans
Perceived Bullsh*tness (e.g., the task feels pointless)	Lawyers	Prepare, draft, and review legal documents, such as wills, deeds, patent applications, mortgages, leases, and contracts.	0.00 (Strongly disagree)	1.93 (Neutral)	*LM*: Essential, substantive, impacts clients' rights, central to organizational goals. *Humans*: Moderate meaningfulness, some bureaucratic perception.
Perceived Value (e.g., I have the freedom to decide how to carry out this task)	Poets, Lyricists, Creative Writers	Plan project arrangements or outlines, and organize material accordingly.	3.80 (Strongly agree)	2.00 (Neutral)	*LM*: Tangible outcomes, autonomy, team contribution. *Humans*: Recognize value but lower perceived impact.
Status Maintenance (e.g., This task helps reinforce my standing in the organization)	General & Operations Managers	Set prices or credit terms for goods or services, based on forecasts of customer demand.	3.83 (Strongly agree)	1.89 (Neutral)	*LM*: Visibility, authority, organizational standing. *Humans*: Less tied to status perception.
EPOCH (e.g., This task requires recognizing and responding appropriately to the emotions of others)	Survey Researchers	Review, classify, and record survey data in preparation for computer analysis.	0.40 (Strongly disagree)	3.20 (Agree)	*LM*: Technical, routine, procedural, minimal emotional/moral impact. *Humans*: Requires some novel ideas and sometimes requires emotional and moral judgments.
Human Flourishing (e.g., This task helps me build or strengthen relationships with colleagues or clients)	HR Specialists	Schedule or conduct new employee orientations.	3.50 (Agree)	1.69 (Neutral)	*LM*: Purpose, social connection, growth-oriented. *Humans*: Perceived as routine or standard procedure.
Psychological Traits of AI Behavior (e.g., Show warmth and care rather than remain neutral and business-like)	Fundraisers	Recruit sponsors, participants, or volunteers for fundraising events.	3.08 (Agree)	0.94 (Disagree)	*LM*: Emotional engagement, tailored reasoning, insight, creativity. *Humans*: Functional, minimal AI-alignment traits perceived.

Table 11. RQ1: Significant task dimensions (FDR < 0.05 and $|\Delta| \geq 0.10$). Estimates from nested RE models; Δ is the back-transformed Likert difference (likely − not-likely) where (+) means that the dimension was more exposed to AI augmentation and (-) it was less exposed to AI augmentation

Task Characteristic	β (95% CI)	Δ (95% CI)
Novel ideas/creativity (+)	0.22698 [0.17405, 0.27991]	0.29298 [0.22466, 0.36131]
Happy/Positive (+)	0.19965 [0.13789, 0.26143]	0.12675 [0.08753, 0.16596]
Freedom in how to do it (+)	0.18587 [0.12659, 0.24516]	0.11487 [0.07823, 0.15151]
In-person interaction (-)	−0.17564 [−0.23024, −0.12035]	−0.23429 [−0.30713, −0.16145]
Emotional Awareness (-)	−0.13643 [−0.19021, −0.08259]	−0.20635 [−0.28781, −0.12492]
Build relationships (-)	−0.13422 [−0.19249, −0.07594]	−0.15360 [−0.22028, −0.08692]
Socially supported (-)	−0.10393 [−0.16434, −0.04352]	−0.10683 [−0.16893, −0.04473]

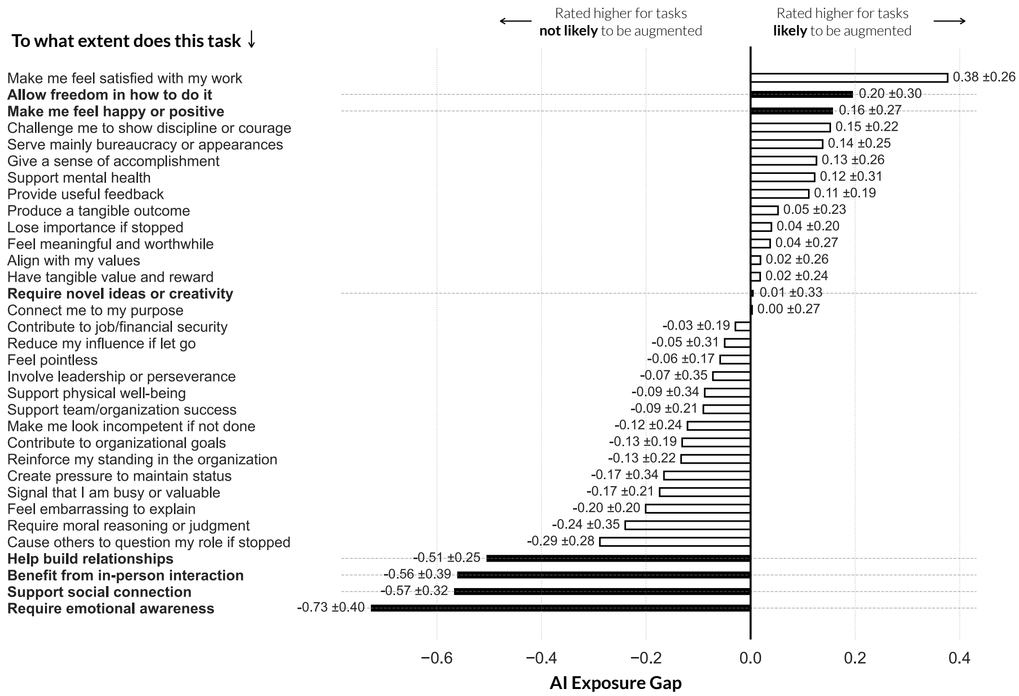


Fig. 9. *RQ1:* AI Exposure Gap by dimensions of meaningful work (rows) based on small-scale human ratings for 171 tasks. A higher gap indicates that a dimension is more strongly associated with tasks likely to be augmented by AI. The gap is computed as the difference in the perceived importance of a dimension between two groups of tasks: those more likely and those less likely to be augmented. We estimate the gaps and 95% confidence intervals by computing average differences for human ratings. Bold names and corresponding black bars indicate those statistically significant different dimensions of meaningful work identified with *LM annotations* based model (Figure 3). Results based on human ratings are highly consistent with LM annotations: tasks rated as likely to be augmented by AI tend to involve satisfaction with work, happiness, and autonomy, whereas tasks rated as not likely to be augmented tend to involve emotional awareness, in-person interaction, relationship building, and social support. One major exception is "Require novel ideas or creativity" where human annotators tended to underestimate the novelty of tasks, whereas LMs captured more nuanced distinctions.

Table 12. Top sectors and exemplar tasks for each significant survey item in RQ1. For each item we select the three sectors with the highest normalized sector mean (z-score) within the relevant subset (likely vs. not likely to be automated). Within each sector we show one exemplar task from the 99th percentile of that item's ratings (fallback: sector max).

Task Characteristic	Exemplar task (Occupation | Sector)
Novel ideas/creativity (+)	Formulate basic layout design or presentation approach and specify material details, such as style and size of type … (Art Directors | Arts, Design, Entertainment, Sports, and Media) Develop, present, or respond to proposals for specific customer requirements, including request for proposal responses … (Sales Engineers | Sales and Related) Prepare scale drawings or architectural designs, using computer-aided design or other tools. (Architects, Except Landscape and Naval | Architecture and Engineering)
Happy/Positive (+)	Modify treatment plans to comply with changes in client status. (Substance Abuse and Behavioral Disorder Counselors | Community and Social Service) Present lectures and conduct discussions to increase students' knowledge and competence using visual aids, such as … (Career/Technical Education Teachers, Postsecondary | Educational Instruction and Library) Create custom illustrations or other graphic elements. (Art Directors | Arts, Design, Entertainment, Sports, and Media)
Freedom in how to do it (+)	Formulate basic layout design or presentation approach and specify material details, such as style and size of type … (Art Directors | Arts, Design, Entertainment, Sports, and Media) Develop or execute strategies to address issues such as energy use, resource conservation, recycling, pollution … (Chief Sustainability Officers | Management) Construct probability tables for events such as fires, natural disasters, and unemployment … (Actuaries | Computer and Mathematical)
Emotional awareness (-)	Counsel individuals or groups to help them understand and overcome personal, social, or behavioral problems affecting … (Educational, Guidance, and Career Counselors and Advisors | Community and Social Service) Perform surgery to prepare the mouth for dental implants and to aid in the regeneration of deficient bone and gum … (Oral and Maxillofacial Surgeons | Healthcare Practitioners and Technical) Provide assistance to the public, such as directions to court offices. (Bailiffs | Protective Service)
In-person interaction (-)	Provide assistance to the public, such as directions to court offices. (Bailiffs | Protective Service) Perform surgery to prepare the mouth for dental implants and to aid in the regeneration of deficient bone and gum … (Oral and Maxillofacial Surgeons | Healthcare Practitioners and Technical) Counsel individuals or groups to help them understand and overcome personal, social, or behavioral problems affecting … (Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)
Build relationships (-)	Counsel individuals or groups to help them understand and overcome personal, social, or behavioral problems affecting … (Educational, Guidance, and Career Counselors and Advisors | Community and Social Service) Present purchase offers to sellers for consideration. (Real Estate Sales Agents | Sales and Related) Mentor new faculty members. (Social Work Teachers, Postsecondary | Educational Instruction and Library)
Socially supported (-)	Counsel individuals or groups to help them understand and overcome personal, social, or behavioral problems affecting … (Educational, Guidance, and Career Counselors and Advisors | Community and Social Service) Mentor new faculty members. (Social Work Teachers, Postsecondary | Educational Instruction and Library) Plan, organize, and conduct occupational therapy programs in hospital, institutional, or community settings to help … (Occupational Therapists | Healthcare Practitioners and Technical)

Table 13. K-means clustering ($k = 10$) of tasks in the 99th percentile for 'novel ideas/creativity,' restricted to tasks judged likely to be augmented. Cluster labels were assigned using GPT-4o, and exemplar tasks come from sectors with the strongest heatmap z-scores.

Cluster (size)	**Label**	**Exemplar task (Occupation | Sector)**
2 (n=148)	Automated Task Management	Plan or coordinate investigation and resolution of customers' reports of technical problems with aircraft or aerospace vehicles. *(Aerospace Engineers | Architecture and Engineering)*
6 (n=140)	Policy Development	Conduct educational programs that provide farmers or farm cooperative members with information that can help them improve agricultural prod… *(Agricultural Engineers | Architecture and Engineering)*
5 (n=128)	Innovative Solutions Evaluation	Analyze project requests, proposals, or engineering data to determine feasibility, productibility, cost, or production time of aerospace or… *(Aerospace Engineers | Architecture and Engineering)*
3 (n=124)	Sustainable Project Planning	Design environmentally sound structural upgrades to existing buildings, such as natural lighting systems, green roofs, or rainwater collect… *(Architects, Except Landscape and Naval | Architecture and Engineering)*
8 (n=112)	Creative Engineering Tasks	Plan or conduct experimental, environmental, operational, or stress tests on models or prototypes of aircraft or aerospace systems or equip… *(Aerospace Engineers | Architecture and Engineering)*
9 (n=107)	Design and Planning Tasks	Prepare scale drawings or architectural designs, using computer-aided design or other tools. *(Architects, Except Landscape and Naval | Architecture and Engineering)*
0 (n=96)	Logistics and Market Analysis	Analyze new medical procedures to forecast likely outcomes. *(Bioengineers and Biomedical Engineers | Architecture and Engineering)*
4 (n=95)	Biofuels Research and Development	Design sensing, measuring, and recording devices, and other instrumentation used to study plant or animal life. *(Agricultural Engineers | Architecture and Engineering)*
7 (n=81)	Creative Visual Production	Create custom illustrations or other graphic elements. *(Art Directors | Arts, Design, Entertainment, Sports, and Media)*
1 (n=79)	Marketing and Promotion	Confer with customers to assess customer needs or obtain feedback. *(Craft Artists | Arts, Design, Entertainment, Sports, and Media)*

Table 14. K-means clustering ($k = 10$) of tasks in the 99th percentile for 'happy/positive' restricted to tasks judged likely to be augmented. Cluster labels were assigned using GPT-4o, and exemplar tasks come from sectors with the strongest heatmap z-scores.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
6 (n=124)	Client Support and Training	Plan and promote career and employment-related programs and events, such as career planning presentations, work experience programs, job fa… *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*
7 (n=79)	Positive Media Production	Use computers, audio-visual aids, and other equipment and materials to supplement presentations. *(Kindergarten Teachers, Except Special Education | Educational Instruction and Library)*
9 (n=72)	Patient Eligibility Assessment	Arrange for medical, psychiatric, and other tests that may disclose causes of difficulties and indicate remedial measures. *(Child, Family, and School Social Workers | Community and Social Service)*
5 (n=70)	Client Rehabilitation Plans	Modify treatment plans to comply with changes in client status. *(Substance Abuse and Behavioral Disorder Counselors | Community and Social Service)*
1 (n=50)	Training and Development	Evaluate students' or individuals' abilities, interests, and personality characteristics, using tests, records, interviews, or professional… *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*
4 (n=44)	Robotics and Mechatronics	Design advanced precision equipment for accurate or controlled applications. *(Mechatronics Engineers | Architecture and Engineering)*
0 (n=41)	Emergency Response Tasks	Provide assistive devices, supportive technology, or assistance accessing facilities, such as restrooms. *(Special Education Teachers, Preschool | Educational Instruction and Library)*
3 (n=37)	Creative Performance Tasks	Portray and interpret roles, using speech, gestures, and body movements, to entertain, inform, or instruct radio, film, television, or live… *(Actors | Arts, Design, Entertainment, Sports, and Media)*
2 (n=35)	Biofuels Innovation	Propose new biofuels products, processes, technologies or applications based on findings from applied biofuels or biomass research projects. *(Biofuels/Biodiesel Technology and Product Development Managers | Management)*
8 (n=28)	Speech and Language Therapy	Evaluate hearing or speech and language test results, barium swallow results, or medical or background information to diagnose and plan tre… *(Speech-Language Pathologists | Healthcare Practitioners and Technical)*

Table 15. K-means clustering ($k = 10$) of tasks in the 99th percentile for 'giving workers freedom and agency' restricted to tasks judged likely to be augmented. Cluster labels were assigned using GPT-4o, and exemplar tasks come from sectors with the strongest heatmap z-scores.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
3 (n=188)	Automated Security Systems	Develop computer information resources, providing for data security and control, strategic computing, and disaster recovery. *(Computer and Information Systems Managers | Management)*
4 (n=141)	Marketing and Strategy Evaluation	Identify, develop, or evaluate marketing strategy, based on knowledge of establishment objectives, market characteristics, and cost and mar… *(Marketing Managers | Management)*
0 (n=136)	Sustainability Strategies	Develop or execute strategies to address issues such as energy use, resource conservation, recycling, pollution reduction, waste eliminatio… *(Chief Sustainability Officers | Management)*
7 (n=124)	Engineering Design Tasks	Identify opportunities to improve plant electrical equipment, controls, or process control methodologies. *(Geothermal Production Managers | Management)*
1 (n=123)	Health Program Management	Maintain awareness of advances in medicine, computerized diagnostic and treatment equipment, data processing technology, government regulat… *(Medical and Health Services Managers | Management)*
6 (n=102)	Design and Layout Tasks	Plan store layouts or design displays. *(General and Operations Managers | Management)*
5 (n=71)	Environmental Planning Tasks	Manage site assessments or environmental studies for wind fields. *(Wind Energy Development Managers | Management)*
2 (n=71)	Marketing and Promotion Tasks	Develop or implement product-marketing strategies, including advertising campaigns or sales promotions. *(General and Operations Managers | Management)*
9 (n=69)	Biofuels Research and Development	Design or conduct applied biodiesel or biofuels research projects on topics, such as transport, thermodynamics, mixing, filtration, distill… *(Biofuels/Biodiesel Technology and Product Development Managers | Management)*
8 (n=44)	Training Program Development	Evaluate instructor performance and the effectiveness of training programs, providing recommendations for improvement. *(Training and Development Managers | Management)*

Table 16. K-means clustering (k = 10) of tasks in the 99th percentile for 'requires emotional awareness' restricted to tasks judged likely to be augmented. Cluster labels were assigned using GPT-4o, and exemplar tasks come from sectors with the strongest heatmap z-scores.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
2 (n=185)	Emotion Awareness in Leadership	Testify in depositions or trials as an expert witness. *(Physicians, Pathologists | Healthcare Practitioners and Technical)*
7 (n=171)	Crisis Intervention and Support	Perform crisis interventions to help ensure the safety of the patients and others. *(Mental Health Counselors | Community and Social Service)*
1 (n=165)	Emotion Awareness Tasks	Provide students with disabilities with assistive devices, supportive technology, and assistance accessing facilities, such as restrooms. *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*
3 (n=140)	Emotion Awareness Tasks	Perform surgery to prepare the mouth for dental implants and to aid in the regeneration of deficient bone and gum tissues. *(Oral and Maxillofacial Surgeons | Healthcare Practitioners and Technical)*
6 (n=129)	Personnel Management	Direct the operations of short stay or specialty units. *(Hospitalists | Healthcare Practitioners and Technical)*
8 (n=115)	Emotion Awareness Tasks	Attend meetings, educational conferences, and training workshops, and serve on committees. *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*
4 (n=105)	Networking and Representation	Deliver presentations to lay or professional audiences. *(Preventive Medicine Physicians | Healthcare Practitioners and Technical)*
0 (n=104)	Emotion Awareness Counseling	Counsel individuals or groups to help them understand and overcome personal, social, or behavioral problems affecting their educational or … *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*
5 (n=60)	Professional Development	Teach pharmacy students serving as interns in preparation for their graduation or licensure. *(Pharmacists | Healthcare Practitioners and Technical)*
9 (n=40)	Student Behavior Management	Establish and enforce administration policies and rules governing student behavior. *(Educational, Guidance, and Career Counselors and Advisors | Community and Social Service)*

Table 17. *RQ2:* Worker–developer misalignment by AI traits with human ratings on 171 tasks. Misalignment is defined as the average absolute difference between worker and developer ratings (Q34–Q45) of the traits they believe AI systems should possess when augmenting tasks. Differences (Δ) are calculated as worker minus developer ratings, with the magnitude ($|\Delta|$) reflecting the size of the misalignment. Reported values aggregate over sectors and traits are grouped into high, mixed, or aligned categories based on percentile thresholds of average absolute misalignment. The bolded traits are those classified into the same categories based on human and LM ratings (Table 18) on the same set of 171 tasks. As with LM rating–based results, the highest misalignments in human ratings occur for Explainable vs Fast/automatic, Straightforward vs. Polite, and Emotional vs. Apathetic. However, for the for the highly misaligned traits, discrepancies emerge: Handle complex vs. Routine work are highly misaligned in human ratings but fall only into mixed misalignment categories in LM ratings. After manual inspection, we found that this is partly due to bias in the small-scale human ratings data, which are more concentrated in certain sectors/occupations.

Trait	$\mu\|\Delta\|$
High misalignment	
(Q36) Explainable vs. Fast/automatic	1.367
(Q40) Straightforward vs. Polite	1.211
(Q34) Handle complex vs. Routine work	1.165
(Q35) Address emotions vs. Apathetic	1.165
Mixed misalignment	
(Q43) Flexible vs. Determined	1.0917
(Q39) Business-like vs. Warm/caring	1.073
(Q41) Tolerant/Open-minded vs. Strict	1.064
(Q38) Generalized vs. Personalized	0.936
Aligned	
(Q42) Precise vs. Simple	0.936
(Q45) Practical vs. Imaginative	0.872
(Q37) Definitive vs. Open to challenge	0.853
(Q44) Simple vs. Comprehensive	0.817

Table 18. *RQ2:* Worker–developer misalignment by AI traits with LM-simulated ratings on 171 tasks. The bolded traits are those classified into the same categories based on human and LM ratings (Table 17) on the same set of 171 tasks.

Trait	$\mu\lvert\Delta\rvert$
High misalignment	
(Q40) Straightforward vs. Polite	1.936
(Q36) Explainable vs. Fast/automatic	0.780
(Q35) Address emotions vs. Apathetic	0.706
(Q42) Precise vs. Simple	0.670
Mixed misalignment	
(Q41) Tolerant/Open-minded vs. Strict	0.6514
(Q39) Business-like vs. Warm/caring	0.523
(Q34) Handle complex vs. Routine work	0.450
(Q43) Flexible vs. Determined	0.450
Aligned	
(Q45) Practical vs. Imaginative	0.404
(Q37) Definitive vs. Open to challenge	0.330
(Q38) Generalized vs. Personalized	0.275
(Q44) Simple vs. Comprehensive	0.110

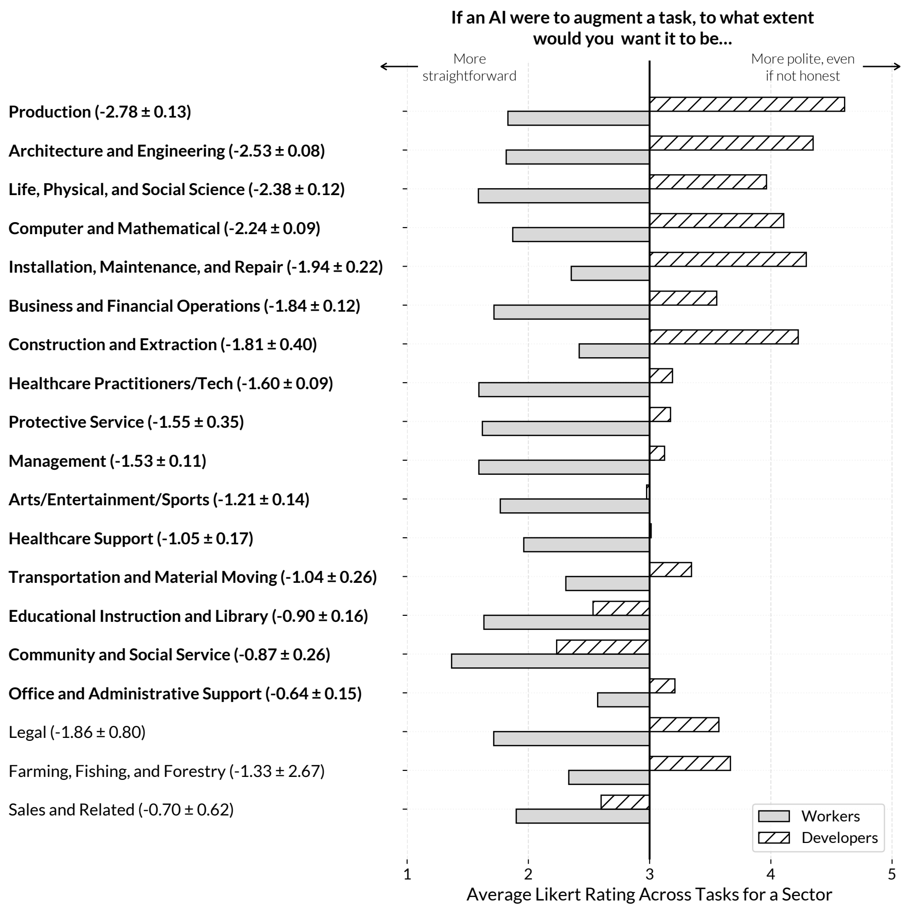


Fig. 10. Worker vs. developer preferences for AI to be straightforward or polite, by sector. Paired bars show mean Likert ratings for each group. Greater distance between bars indicates stronger misalignment; misalignment scores with standard errors are shown on the y-axis.

Table 19. Clusters of tasks with the lowest worker–developer misalignment on the trait definitive vs. open to challenge, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
0 (n=263)	Strategic Decision Making	Investigate traffic problems and recommend methods to improve traffic flow or safety. *(Transportation Engineers | Architecture and Engineering)*
3 (n=254)	Strategic Evaluation and Analysis	Analyze data on conditions such as site location, drainage, or structure location for environmental reports or landscaping plans. *(Landscape Architects | Architecture and Engineering)*
5 (n=217)	Flexible Decision-Making	Analyze new medical procedures to forecast likely outcomes. *(Bioengineers and Biomedical Engineers | Architecture and Engineering)*
1 (n=212)	Flexible Decision-Making	Develop or assist in the development of transportation-related computer software or computer processes. *(Transportation Engineers | Architecture and Engineering)*
2 (n=187)	Research and Development Planning	Prepare scale drawings or architectural designs, using computer-aided design or other tools. *(Architects, Except Landscape and Naval | Architecture and Engineering)*
7 (n=168)	Task Coordination and Improvement	Develop processes to separate components of liquids or gases or generate electrical currents, using controlled chemical processes. *(Chemical Engineers | Architecture and Engineering)*
9 (n=158)	Flexible Decision Making	Document equipment or process details of radio frequency identification device (RFID) technology. *(Radio Frequency Identification Device Specialists | Architecture and Engineering)*
6 (n=132)	Quality Control Analysis	Plan or conduct experimental, environmental, operational, or stress tests on models or prototypes of aircraft or aerospace systems or equip... *(Aerospace Engineers | Architecture and Engineering)*
8 (n=110)	Data Analysis and Review	Design sensing, measuring, and recording devices, and other instrumentation used to study plant or animal life. *(Agricultural Engineers | Architecture and Engineering)*
4 (n=89)	Task Evaluation and Development	Train users in task techniques or ergonomic principles. *(Human Factors Engineers and Ergonomists | Architecture and Engineering)*

Table 20. Clusters of tasks with the lowest worker–developer misalignment on the trait generalization vs. personalized, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	Cluster label	Exemplar task (Occupation | Sector)
9 (n=278)	Financial Coordination	Prepare responses to customer requests for information, such as product data, written regulatory affairs statements, surveys, or questionna… *(Regulatory Affairs Specialists | Business and Financial Operations)*
6 (n=255)	Quality and Compliance Monitoring	Examine damaged vehicle to determine extent of structural, body, mechanical, electrical, or interior damage. *(Insurance Appraisers, Auto Damage | Business and Financial Operations)*
0 (n=240)	Technical Project Management	Develop and implement technical project management tools, such as plans, schedules, and responsibility and compliance matrices. *(Logisticians | Business and Financial Operations)*
5 (n=231)	Medical Task Execution	Perform medicolegal examinations and autopsies, conducting preliminary examinations of the body to identify victims, locate signs of trauma… *(Coroners | Business and Financial Operations)*
3 (n=211)	Health Program Management	Prepare reports of findings, illustrating data graphically and translating complex findings into written text. *(Market Research Analysts and Marketing Specialists | Business and Financial Operations)*
1 (n=205)	Marketing Strategy Development	Verify and analyze data used in settling claims to ensure that claims are valid and that settlements are made according to company practice… *(Claims Adjusters, Examiners, and Investigators | Business and Financial Operations)*
7 (n=171)	Design and Media Tasks	Compose images of products, using video or still cameras, lighting equipment, props, or photo or video editing software. *(Online Merchants | Business and Financial Operations)*
2 (n=165)	Energy and Resource Management	Evaluate the use of technologies, such as global positioning systems (GPS), radio-frequency identification (RFID), route navigation softwar… *(Logistics Engineers | Business and Financial Operations)*
8 (n=139)	Communication Systems Management	Investigate, evaluate, and settle claims, applying technical knowledge and human relations skills to effect fair and prompt disposal of cas… *(Claims Adjusters, Examiners, and Investigators | Business and Financial Operations)*
4 (n=96)	Training Program Evaluation	Obtain, organize, or develop training procedure manuals, guides, or course materials, such as handouts or visual materials. *(Training and Development Specialists | Business and Financial Operations)*

Table 21. Clusters of tasks with the lowest worker–developer misalignment on the trait simple vs. deeply insightful/comprehensive, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	Cluster label	Exemplar task (Occupation | Sector)
2 (n=410)	Health Program Management	Monitor patients' performance in therapy activities, providing encouragement. *(Occupational Therapy Assistants | Healthcare Support)*
0 (n=394)	Strategic Operations Management	Work under the direction of occupational therapists to plan, implement, or administer educational, vocational, or recreational programs tha… *(Occupational Therapy Assistants | Healthcare Support)*
9 (n=360)	Task Management and Support	Prepare, maintain, and record records of inventories, receipts, purchases, or deliveries, using a variety of computer screen formats. *(Pharmacy Aides | Healthcare Support)*
4 (n=339)	Sales Promotion Analysis	Collect and compile data to document clients' performance or assess program quality. *(Speech-Language Pathology Assistants | Healthcare Support)*
8 (n=284)	Biofuels Research Tasks	Fabricate and fit orthodontic appliances and materials for patients, such as retainers, wires, or bands. *(Dental Assistants | Healthcare Support)*
3 (n=275)	Quality Control Tasks	Consult with managers or other personnel to resolve problems in areas such as equipment performance, output quality, or work schedules. *(First-Line Supervisors of Office and Administrative Support Workers | Office and Administrative Support)*
6 (n=259)	Equipment Maintenance Coordination	Design, fabricate, or repair assistive devices or make adaptive changes to equipment or environments. *(Occupational Therapy Assistants | Healthcare Support)*
5 (n=229)	Evaluate Training Effectiveness	Select or prepare speech-language instructional materials. *(Speech-Language Pathology Assistants | Healthcare Support)*
1 (n=165)	Environmental Monitoring	Read and effectively interpret small-scale maps and information from a computer screen to determine locations and provide directions. *(Public Safety Telecommunicators | Office and Administrative Support)*
7 (n=150)	Biofuels Data Analysis	Expose dental diagnostic x-rays. *(Dental Assistants | Healthcare Support)*

Table 22. Clusters of tasks with the highest worker–developer misalignment on the trait straightforward vs. polite even if not honest, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
8 (n=135)	Quality Control Oversight	Plan or conduct experimental, environmental, operational, or stress tests on models or prototypes of aircraft or aerospace systems or equip… *(Aerospace Engineers | Architecture and Engineering)*
0 (n=131)	Technical Design Tasks	Create three-dimensional or interactive representations of designs, using computer-assisted design software. *(Architects, Except Landscape and Naval | Architecture and Engineering)*
1 (n=130)	Task Coordination and Documentation	Prepare documentation containing information such as confidential descriptions or specifications of proprietary hardware or software, produ… *(Electronics Engineers, Except Computer | Architecture and Engineering)*
6 (n=129)	Data Analysis and Forecasting	Determine usefulness of new radio frequency identification device (RFID) technologies. *(Radio Frequency Identification Device Specialists | Architecture and Engineering)*
9 (n=128)	Technical Task Management	Develop or assist in the development of transportation-related computer software or computer processes. *(Transportation Engineers | Architecture and Engineering)*
4 (n=107)	Data Analysis Tasks	Store, retrieve, and manipulate data for analysis of system capabilities and requirements. *(Computer Hardware Engineers | Architecture and Engineering)*
3 (n=103)	Technical Oversight Tasks	Inspect completed installations and observe operations to ensure conformance to design and equipment specifications and compliance with ope… *(Electrical Engineers | Architecture and Engineering)*
2 (n=86)	Complex Technical Tasks	Conduct research related to a range of nanotechnology topics, such as packaging, heat transfer, fluorescence detection, nanoparticle disper… *(Nanosystems Engineers | Architecture and Engineering)*
5 (n=78)	Technical Surveying Tasks	Prepare and alter trace maps, charts, tables, detailed drawings, and three-dimensional optical models of terrain using stereoscopic plottin… *(Cartographers and Photogrammetrists | Architecture and Engineering)*
7 (n=77)	Task Monitoring and Reporting	Analyze new medical procedures to forecast likely outcomes. *(Bioengineers and Biomedical Engineers | Architecture and Engineering)*

Table 23. Clusters of tasks with the highest worker–developer misalignment on the trait tolerant/open-minded vs. strict, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	**Cluster label**	**Exemplar task (Occupation | Sector)**
2 (n=20)	Strict Process Improvement	Identify opportunities to improve plant electrical equipment, controls, or process control methodologies. *(Geothermal Production Managers | Management)*
0 (n=15)	Strict Monitoring and Planning	Monitor food preparation methods, portion sizes, and garnishing and presentation of food to ensure that food is prepared and presented in a… *(Food Service Managers | Management)*
4 (n=11)	Data Analysis and Evaluation	Collect and analyze survey data, regulatory information, and data on demographic and employment trends to forecast enrollment patterns and … *(Education Administrators, Kindergarten through Secondary | Management)*
8 (n=7)	Strict Budget Management	Develop or review budgets, annual plans, power contracts, power rates, standing operating procedures, power reviews, or engineering studies. *(Hydroelectric Production Managers | Management)*
1 (n=6)	Research and Evaluation Tasks	Conduct research to develop methodologies, instrumentation, and procedures for medical application, analyzing data and presenting findings. *(Epidemiologists | Life, Physical, and Social Science)*
9 (n=5)	Data Analysis and Marketing	Monitor and analyze sales promotion results to determine cost effectiveness of promotion campaigns. *(Advertising and Promotions Managers | Management)*
5 (n=5)	Strict Risk Management	Create scenarios to reestablish operations from various types of business disruptions. *(Business Continuity Planners | Business and Financial Operations)*
3 (n=5)	Environmental Stewardship Analysis	Provide for stewardship of plant or animal resources or habitats, studying land use, monitoring animal populations, or providing shelter, r… *(Natural Sciences Managers | Management)*
6 (n=4)	Strict Supply Chain Oversight	Establish or monitor specific supply chain-based performance measurement systems. *(Transportation, Storage, and Distribution Managers | Management)*
7 (n=4)	Rigid Web Development Guidelines	Create Web models or prototypes that include physical, interface, logical, or data models. *(Web Developers | Computer and Mathematical)*

Table 24. Clusters of tasks with the highest worker–developer misalignment on the trait practical vs. imaginative, identified using MPNet embeddings and K-means clustering. Cluster labels were generated using GPT-4o.

Cluster (size)	Cluster label	Exemplar task (Occupation | Sector)
8 (n=23)	Highly Structured Tasks	Plan sequences of operations, applying knowledge of physical properties of workpiece materials. *(Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic | Production)*
1 (n=17)	Technical Operations	Monitor power plant equipment and indicators to detect evidence of operating problems. *(Power Plant Operators | Production)*
3 (n=10)	Highly Practical Tasks	Study traffic delays by noting times of delays, the numbers of vehicles affected, and vehicle speed through the delay area. *(Traffic Technicians | Transportation and Material Moving)*
6 (n=8)	Routine Equipment Monitoring	Operate or maintain distributed power generation equipment, including fuel cells or microturbines, to produce energy on-site for manufactur… *(Power Plant Operators | Production)*
2 (n=8)	Highly technical lab tasks	Perform laboratory procedures following protocols including deoxyribonucleic acid (DNA) sequencing, cloning and extraction, ribonucleic aci… *(Molecular and Cellular Biologists | Life, Physical, and Social Science)*
4 (n=6)	Administrative Support Tasks	Maintain databases, mailing lists, telephone networks, and other information to facilitate the functioning of health education programs. *(Health Education Specialists | Community and Social Service)*
5 (n=5)	Data Management Tasks	Enter computer commands to store or retrieve parts patterns, graphic displays, or programs that transfer data to other media. *(Computer Numerically Controlled Tool Programmers | Production)*
7 (n=3)	Technical Calibration Tasks	Pretest and calibrate anesthesia delivery systems and monitors. *(Anesthesiologist Assistants | Healthcare Practitioners and Technical)*
9 (n=1)	Technical Source Selection	Select sources from which programming will be received or through which programming will be transmitted. *(Broadcast Technicians | Arts, Design, Entertainment, Sports, and Media)*
0 (n=1)	Strict Compliance Tasks	Check building codes and zoning bylaws to determine any effects on the properties being appraised. *(Appraisers and Assessors of Real Estate | Business and Financial Operations)*

Table 25. Sector-level misalignment for Q37 (Final decisions < Sometimes flexible < Open to being challenged). Misalignment ($\Delta_{t,q}$) is calculated as worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred systems to be more open to challenge than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95% confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Installation, Maintenance, and Repair	**-0.6**	**0.072**	**[-0.744, -0.456]**	**65**
Architecture and Engineering	**-0.567**	**0.024**	**[-0.614, -0.520]**	**476**
Transportation and Material Moving	**-0.545**	**0.077**	**[-0.700, -0.391]**	**55**
Healthcare Practitioners and Technical	-0.48	0.03	[-0.538, -0.421]	392
Production	-0.476	0.056	[-0.586, -0.366]	166
Healthcare Support	-0.43	0.064	[-0.558, -0.303]	79
Management	-0.396	0.036	[-0.468, -0.325]	217
Computer and Mathematical	-0.392	0.021	[-0.434, -0.350]	574
Construction and Extraction	-0.387	0.11	[-0.613, -0.161]	31
Protective Service	-0.379	0.092	[-0.567, -0.191]	29
Office and Administrative Support	-0.379	0.064	[-0.505, -0.253]	182
Life, Physical, and Social Science	-0.376	0.029	[-0.434, -0.318]	279
Community and Social Service	-0.333	0.088	[-0.512, -0.154]	30
Farming, Fishing, and Forestry	-0.333	0.333	[-1.768, 1.101]	3
Business and Financial Operations	-0.325	0.029	[-0.383, -0.268]	289
Arts, Design, Entertainment, Sports, and Media	-0.261	0.032	[-0.325, -0.197]	203
Sales and Related	-0.2	0.2	[-0.652, 0.252]	10
Legal	-0.143	0.261	[-0.781, 0.495]	7
Educational Instruction and Library	-0.122	0.041	[-0.204, -0.040]	90

3017 Table 26. Sector-level misalignment for Q38 (Generalization < Some adjustment < Personalized). Misalignment ($\Delta_{t,q}$) is calculated as
3018 worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred more personalized
3019 systems than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95%
3020 confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Life, Physical, and Social Science	-0.43	0.069	[-0.565, -0.295]	279
Business and Financial Operations	-0.367	0.054	[-0.474, -0.260]	289
Healthcare Practitioners and Technical	-0.301	0.047	[-0.393, -0.209]	392
Construction and Extraction	-0.29	0.141	[-0.577, -0.003]	31
Installation, Maintenance, and Repair	-0.246	0.17	[-0.586, 0.093]	65
Office and Administrative Support	-0.231	0.08	[-0.389, -0.072]	182
Sales and Related	-0.2	0.133	[-0.502, 0.102]	10
Computer and Mathematical	-0.199	0.048	[-0.292, -0.105]	574
Architecture and Engineering	-0.189	0.057	[-0.302, -0.076]	476
Legal	-0.143	0.261	[-0.781, 0.495]	7
Protective Service	-0.103	0.091	[-0.289, 0.082]	29
Production	-0.102	0.086	[-0.272, 0.068]	166
Management	-0.069	0.052	[-0.172, 0.034]	217
Community and Social Service	0.067	0.067	[-0.070, 0.203]	30
Healthcare Support	0.051	0.09	[-0.128, 0.230]	79
Transportation and Material Moving	0.036	0.142	[-0.248, 0.321]	55
Arts, Design, Entertainment, Sports, and Media	0.02	0.048	[-0.075, 0.114]	203
Educational Instruction and Library	0.011	0.04	[-0.069, 0.091]	90
Farming, Fishing, and Forestry	0.0	1.0	[-4.303, 4.303]	3

Table 27. Sector-level misalignment for Q40 (Straightforward < Neutral < Polite even if not honest). Misalignment ($\Delta_{t,q}$) is calculated as worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred systems to be more polite (even at the expense of honesty) than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95% confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Production	**-2.777**	**0.135**	**[-3.044, -2.511]**	**166**
Architecture and Engineering	**-2.532**	**0.085**	**[-2.698, -2.365]**	**476**
Life, Physical, and Social Science	**-2.376**	**0.116**	**[-2.604, -2.149]**	**279**
Computer and Mathematical	**-2.235**	**0.086**	**[-2.404, -2.067]**	**574**
Installation, Maintenance, and Repair	**-1.938**	**0.218**	**[-2.374, -1.503]**	**65**
Business and Financial Operations	**-1.837**	**0.118**	**[-2.069, -1.605]**	**289**
Construction and Extraction	**-1.806**	**0.403**	**[-2.629, -0.984]**	**31**
Healthcare Practitioners and Technical	**-1.597**	**0.088**	**[-1.770, -1.424]**	**392**
Protective Service	**-1.552**	**0.353**	**[-2.275, -0.829]**	**29**
Management	**-1.525**	**0.113**	**[-1.748, -1.303]**	**217**
Arts, Design, Entertainment, Sports, and Media	**-1.207**	**0.137**	**[-1.476, -0.938]**	**203**
Healthcare Support	**-1.051**	**0.169**	**[-1.387, -0.714]**	**79**
Transportation and Material Moving	**-1.036**	**0.262**	**[-1.562, -0.511]**	**55**
Educational Instruction and Library	**-0.9**	**0.156**	**[-1.211, -0.589]**	**90**
Community and Social Service	**-0.867**	**0.257**	**[-1.392, -0.341]**	**30**
Office and Administrative Support	**-0.637**	**0.154**	**[-0.941, -0.333]**	**182**
Legal	-1.857	0.8	[-3.814, 0.100]	7
Farming, Fishing, and Forestry	-1.333	2.667	[-12.807, 10.140]	3
Sales and Related	-0.7	0.616	[-2.092, 0.692]	10

Table 28. Sector-level misalignment for Q41 (Tolerant/Open-minded < Fair < Strict). Misalignment ($\Delta_{t,q}$) is calculated as worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred stricter systems than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95% confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Community and Social Service	**1.267**	**0.106**	**[1.049, 1.484]**	**30**
Educational Instruction and Library	**1.067**	**0.077**	**[0.913, 1.220]**	**90**
Management	**0.76**	**0.072**	**[0.618, 0.902]**	**217**
Sales and Related	**0.7**	**0.213**	**[0.217, 1.183]**	**10**
Arts, Design, Entertainment, Sports, and Media	**0.655**	**0.059**	**[0.538, 0.772]**	**203**
Healthcare Support	0.443	0.082	[0.280, 0.606]	79
Business and Financial Operations	0.415	0.057	[0.304, 0.527]	289
Healthcare Practitioners and Technical	0.401	0.047	[0.308, 0.493]	392
Life, Physical, and Social Science	0.348	0.052	[0.244, 0.451]	279
Protective Service	-0.276	0.139	[-0.562, 0.010]	29
Computer and Mathematical	0.261	0.036	[0.190, 0.333]	574
Construction and Extraction	0.161	0.105	[-0.053, 0.375]	31
Installation, Maintenance, and Repair	0.154	0.091	[-0.028, 0.336]	65
Office and Administrative Support	0.154	0.049	[0.057, 0.251]	182
Legal	-0.143	0.143	[-0.492, 0.207]	7
Architecture and Engineering	0.044	0.026	[-0.007, 0.095]	476
Production	0.024	0.023	[-0.020, 0.069]	166
Transportation and Material Moving	0.018	0.066	[-0.114, 0.151]	55
Farming, Fishing, and Forestry	0.0	0.0	[0.000, 0.000]	3

Table 29. Sector-level misalignment for Q44 (Simple < Some depth < Deeply insightful/comprehensive). Misalignment ($\Delta_{t,q}$) is calculated as worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred systems that are more deeply insightful and comprehensive than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95% confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Healthcare Support	0.241	0.079	[0.083, 0.398]	79
Transportation and Material Moving	0.2	0.084	[0.032, 0.368]	55
Office and Administrative Support	0.165	0.068	[0.031, 0.298]	182
Installation, Maintenance, and Repair	0.154	0.055	[0.045, 0.263]	65
Protective Service	0.138	0.065	[0.004, 0.271]	29
Production	0.102	0.05	[0.005, 0.200]	166
Construction and Extraction	0.065	0.113	[-0.166, 0.295]	31
Arts, Design, Entertainment, Sports, and Media	0.039	0.024	[-0.008, 0.087]	203
Community and Social Service	0.033	0.033	[-0.035, 0.102]	30
Business and Financial Operations	-0.031	0.019	[-0.069, 0.007]	289
Management	-0.023	0.017	[-0.056, 0.010]	217
Educational Instruction and Library	0.022	0.016	[-0.009, 0.053]	90
Healthcare Practitioners and Technical	0.015	0.019	[-0.022, 0.053]	392
Computer and Mathematical	0.014	0.01	[-0.005, 0.033]	574
Life, Physical, and Social Science	-0.007	0.011	[-0.030, 0.015]	279
Architecture and Engineering	-0.006	0.009	[-0.024, 0.012]	476
Farming, Fishing, and Forestry	0.0	0.577	[-2.484, 2.484]	3
Legal	0.0	0.378	[-0.925, 0.925]	7
Sales and Related	0.0	0.0	[0.000, 0.000]	10

Table 30. Sector-level misalignment for Q45 (Practical < Somewhat creative < Imaginative). Misalignment ($\Delta_{t,q}$) is calculated as worker rating minus developer rating for a given task t and trait q; negative values indicate developers preferred more imaginative systems than workers. Reported values include the average misalignment score for tasks within a sector, its standard error (SE), 95% confidence intervals, and the number of tasks (N) within each sector.

Sector	$\frac{1}{N}\sum_{t=1}^{N}\Delta_{t,q}$	SE	95% CI	N
Production	**-1.09**	**0.077**	**[-1.243, -0.938]**	**166**
Transportation and Material Moving	**-0.6**	**0.146**	**[-0.892, -0.308]**	**55**
Farming, Fishing, and Forestry	-1.333	0.333	[-2.768, 0.101]	3
Office and Administrative Support	-0.467	0.06	[-0.585, -0.349]	182
Healthcare Practitioners and Technical	-0.439	0.051	[-0.539, -0.339]	392
Installation, Maintenance, and Repair	-0.338	0.121	[-0.579, -0.097]	65
Business and Financial Operations	-0.318	0.05	[-0.416, -0.221]	289
Life, Physical, and Social Science	-0.269	0.052	[-0.371, -0.166]	279
Management	-0.23	0.052	[-0.333, -0.128]	217
Construction and Extraction	-0.226	0.165	[-0.563, 0.112]	31
Healthcare Support	-0.203	0.094	[-0.390, -0.015]	79
Arts, Design, Entertainment, Sports, and Media	-0.182	0.047	[-0.275, -0.090]	203
Architecture and Engineering	-0.181	0.039	[-0.258, -0.104]	476
Computer and Mathematical	-0.162	0.034	[-0.230, -0.094]	574
Protective Service	-0.103	0.174	[-0.461, 0.254]	29
Sales and Related	-0.1	0.18	[-0.506, 0.306]	10
Community and Social Service	0.067	0.143	[-0.226, 0.360]	30
Educational Instruction and Library	-0.067	0.044	[-0.154, 0.021]	90
Legal	0.0	0.309	[-0.755, 0.755]	7