The Organizational Design of Human-AI Architectures:

A Task-Based Framework and Evidence from Crowd Science Projects

Abstract

Scholars are making great progress towards understanding AI's impacts on organizational processes and outcomes, but they typically treat AI adoption as an exogenous shock. We complement this work by developing a structural framework to understand when and how organizations deploy AI. Leveraging the task-based model, we theorize how task routinization and physicality predict the use of human-only, automation, and augmentation architectures. We test this framework using unique data on over 1,200 crowd science projects that each involve one well-defined primary task. Our results show that automation is strongly predicted by high task routinization, but likely for different reasons than in the context of prior technologies such as computers. Augmentation (joint human+AI workflows) is most likely for tasks with medium routinization and for "hybrid" tasks that require both cognitive processing and material agency. Our framework suggests that task physicality can serve as a durable structural anchor such that augmentation will remain important even as AI's cognitive capabilities continue to advance. The case for augmentation may be further strengthened when organizations optimize not only task efficiency but also value employee learning and human capital development. Overall, we provide a structural lens to understand diverse trajectories of the future of work and contribute to a deeper understanding of how the general "exposure" to AI plays out within the context of individual organizations.

1 Introduction

The capabilities of artificial intelligence (AI) are advancing rapidly, raising important questions regarding AI's impact on workers and the future of work. Economists and organizational scholars have studied these questions using complementary perspectives. Economists studying AI focus on labor market outcomes such as employment and wages through the lens of substitution and complementarity (Acemoglu & Restrepo, 2019, 2020; Autor, 2024). Although their "task-based model" provides an important micro-foundation by drawing attention to the role of task characteristics, the organization is typically treated as a black box, yielding little insight into the organizational mechanisms that give rise to different observed macro-level outcomes (Autor et al., 2003; Frank et al., 2019).

Organizational scholars, on the other hand, offer rich accounts of the organizational consequences of AI adoption. This includes, among others, research on how professionals navigate algorithmic opacity and judgment when working with AI (Lebovitz et al., 2022; Pachidi et al., 2021), how workers respond to perceived threats of automation and algorithmic management (Kellogg et al., 2020; Möhlmann et al., 2021), and how the adoption of AI affects workers' performance (Allen & Choudhury, 2022; Boussioux et al., 2024; Dell'Acqua et al., 2023; Jia et al., 2023). While organizational scholars have extensively explored the consequences of AI adoption, however, they typically treat the adoption itself as an exogenous shock or experimental treatment rather than a deliberate organizational design choice.

What is missing at the intersection of these two literatures is a structural framework that explains when organizations choose to adopt AI and whether they do so by automating (replacing workers with AI) or by augmenting (integrating AI and workers in joint workflows). We bridge this gap by building on the task-based model (Acemoglu & Autor, 2011; Autor et al., 2003) and the micro-structural approach to organizational design (Puranam, 2018; Puranam

et al., 2014) to theorize the role of task routinization and task physicality in shaping organizations' decisions on whether and how to deploy AI. We resolve conceptual ambiguities by using organizationally-grounded definitions of automation and augmentation that focus on ex-ante task allocation and the integration of task inputs rather than ex-post productivity effects (Choudhary et al., 2025; Engelbart, 1962; Lebovitz et al., 2022; Puranam, 2021). Moreover, we avoid aggregation paradoxes by focusing on the task level of analysis, abstracting from confounding effects that can emerge across tasks or occupations (Acemoglu & Restrepo, 2019; Raisch & Krakowski, 2021). In line with prior work, we define AI as software-based technologies that utilize machine learning to identify patterns and generate predictions from data but that lack material agency (Agrawal et al., 2018; Felten et al., 2021).[1]

We test our predictions regarding the relationships between task characteristics and human-AI architectures in the context of "crowd science", research projects that involve participants in activities such as data collection or creative problem-solving (Haklay et al., 2021; Sauermann & Franzoni, 2015). One well-known example is the project *eBird*, which has contributed roughly half the data in the Global Biodiversity Information Facility (Appendix A lists weblinks to all projects written in italics).[2] Another is the project *Foldit*, which was an important step in David Baker's path towards the 2024 Nobel Prize in Chemistry for his work on protein design.[3] Crowd science is novel and important in itself, but it is also a particularly

[1] We focus on predictive AI that uses task-related training data to make predictions in areas such as image recognition, finding hidden patterns in data, or forecasting future outcomes (Agrawal et al., 2018; Brynjolfsson & Mitchell, 2017). We do not consider generative AI that creates new artifacts based on a general underlying model (e.g., ChatGPT). We did not observe project-level applications of generative AI in our data, suggesting that our focus facilitates theorizing without constraining the empirical analysis. Of course, future research should also study the adoption of generative AI.

[2] https://ebird.org/news/2021-year-in-review.

[3] https://www.nobelprize.org/uploads/2024/12/baker-lecture.pdf.

suitable context for our study because each project typically involves a single primary task, enabling us to measure both task characteristics and AI deployment decisions without the aggregation effects typically found in occupation- or firm-level studies. Collectively, the more than 1,200 projects in our sample span a broad range of tasks that differ with respect to their degree of routinization as well as task physicality.

Our results show that a task-based framework provides a unified logic for predicting whether organizations deploy AI to automate or augment. First, we find that high task routinization serves as a primary predictor of automation – the choice to assign task instances to AI for independent execution. This was not obvious given that modern AI learns from data and does not require the codified instructions that limited prior automation technologies such as robots and computers in non-routine contexts (Autor et al., 2003). Yet, while AI is in principle capable of handling non-routine tasks through advanced pattern recognition (Aghion et al., 2018; Brynjolfsson & Mitchell, 2017; Tschang & Almirall, 2021), low routinization now appears to impose different constraints such as insufficient data and data shifts that impede the reliable execution of task instances by AI (Choudhary et al., 2025). Organizations are also more likely to use AI for automation when tasks are cognitive rather than physical or hybrid, reflecting a key constraint for software-based AI in settings that require material agency.

Moving beyond the prior literature's focus on automation, our results show that task characteristics also systematically predict decisions to deploy AI for augmentation: Organizations are significantly more likely to design joint human-AI architectures for tasks that are of medium routinization and for hybrid tasks. These results suggest that augmentation emerges as a specific organizational design solution for tasks that are sufficiently patterned to benefit from AI's cognitive support, yet too complex or physically tethered for independent AI execution. Our framework suggests a fundamental asymmetry going forward: While rapid improvements in AI capabilities may soon lead organizations to shift cognitive tasks towards

automation regardless of their degree of routinization, the sensorimotor requirements of hybrid tasks are likely to create a durable structural anchor favoring augmentation.[4]

An interesting aspect of our empirical context is that some organizations pursue not only task efficiency (i.e., the production of scientific outputs) but also goals related to the learning and education of their participants. Exploratory analyses show that organizations that do so are less likely to use automation, while they are just as likely to use AI for augmentation. This observation speaks to recent discussions around the impact of automation on human capital development (Beane, 2019; Ellis & Bindley, 2025; Garicano & Rayo, 2025) and encourages scholars to consider what objective function organizations maximize when designing human-AI architectures.

Our study contributes to three major literatures. First, we contribute to organizational scholarship by shifting the focus from the consequences of AI adoption to its determinants. Recent research has provided answers to important questions such as how workers react to AI or how the use of AI for augmentation affects their performance (Dell'Acqua et al., 2023; Jia et al., 2023; Kellogg et al., 2020; Lebovitz et al., 2022). In contrast, we treat AI adoption as an explicit design choice made by organizations and we offer a structural framework to predict whether organizations continue to rely solely on humans, use AI for automation, or use AI in joint human+AI workflows. We also complement prior research that has focused on AI in individual organizations or for specific use cases (Boussioux et al., 2024; Choudhury et al., 2020; Jia et al., 2023) by providing first quantitative evidence on the relationships between task characteristics and human-AI architectures across a large sample of organizations.

[4] Although beyond the scope of our definition of AI, robotics is unlikely to eliminate this asymmetry (see Section 6).

Second, we contribute to research on the economics of AI by studying the organizational implications of the task-based model (Acemoglu & Autor, 2011; Autor et al., 2003). While this model has been instrumental in explaining macro-level impacts of technology on employment or wages, the assumed organizational underpinnings of these shifts are typically difficult to observe. We offer an empirical test of the "first stage" of the task-based model, studying how task characteristics map to organizations' decisions to allocate tasks to humans and/or AI.[5] Our results demonstrate the explanatory power of the task-based model at the micro-level and suggests that joint human-AI workflows constitute a distinct design choice rather than a merely transitional stage on the path towards full automation. Our empirical test of this first stage also contributes to bridging the gap between task characteristics and AI capabilities on the one hand, and aggregate labor market outcomes on the other. This organizational bridge may be important to better understand why studies of the labor market impact of AI show different – and sometimes conflicting – results (Acemoglu et al., 2022; Brynjolfsson et al., 2025; Humlum & Vestergaard, 2025).

Third, we contribute to the growing literature on AI exposure (Acemoglu et al., 2022; Eloundou et al., 2024; Felten et al., 2021; Frey & Osborne, 2017). This literature estimates the potential for AI application in specific tasks and occupations, but it frequently assumes a direct link from exposure to automation or remains agnostic as to how AI will be integrated into organizational workflows. We complement this work by theorizing linkages between task characteristics and organizations' choice to automate versus augment, a choice that has

[5] In addition to stating that the mapping between task characteristics and human vs. machine capabilities will drive firms' allocation of tasks to factors of production (what we call "the first stage"), the task model also speaks to downstream implications for productivity, wages, and the emergence of new tasks (what we might call the "second stage") (Acemoglu et al., 2025; Autor et al., 2003). These second-stage effects are beyond the scope of our paper.

significant implications for how exposure plays out for affected workers (Acemoglu & Restrepo, 2020; Frank et al., 2019). Our study reveals a significant opportunity to make research on AI exposure more powerful by treating automation and augmentation as distinct forms of exposure. Existing methods to estimate exposure likely have the capacity to make this distinction, as they often measure the very task characteristics we identify as structural determinants. Thus, future work could move beyond predicting general AI exposure towards predicting different deployment architectures – automation that replaces humans with AI and augmentation that keeps workers employed by bundling human and AI capabilities.

In section 2, we conceptualize automation and augmentation and develop hypotheses regarding the role of task characteristics in shaping organizations' use of different human-AI architectures. We then provide background on our empirical context in section 3. We discuss data and measures in section 4, followed by the main analysis as well as supplementary analyses and robustness checks in section 5. We conclude by discussing the generalizability of our results, opportunities for future research, as well as our contributions in section 6.

2 Theory and Hypotheses

2.1 Background: Input-based lens and task level of analysis

An important hurdle in studying AI-based automation and augmentation is ambiguity in how these constructs are defined and used across relevant literatures. We argue that this ambiguity arises from two conceptual conflations that hinder consistent theorizing and measurement: the blurring of input- and output-based definitions and the collapsing of different levels of analysis. In the following, we briefly discuss this ambiguity and clarify the foundation for our subsequent theorizing (see Appendix B for an extended discussion).

First, organizational scholars and economists largely agree on the definition of automation as the substitution of machine inputs for human labor inputs in the performance of a task

(Agrawal et al., 2023; Autor et al., 2003; Felten et al., 2021; Raisch & Krakowski, 2021). However, they diverge on the conceptualization of augmentation. Economists typically define augmentation through its ex-post impact on productivity or labor demand – treating a technology as "augmenting" if it increases the marginal product of labor (Acemoglu et al., 2025; Autor et al., 2003). Although useful to model the macro implications of technologies with assumed productivity impacts, this approach remains agnostic as to how concretely AI is deployed in organizations. Organizational scholars are less consistent in their definition of augmentation – while some also assume that augmentation implies higher performance or productivity, others define augmentation based on the configuration of labor and machine inputs, where augmentation means that humans and AI collaborate on the same task (Choudhary et al., 2025; Lebovitz et al., 2022; Puranam, 2021; Raisch & Krakowski, 2021). This input-based definition of augmentation draws attention to organizational design decisions but it also enables scholars to explicitly study when joint human+AI architectures improve performance – and when they do not (Dell'Acqua et al., 2023; Fügener et al., 2026). In this paper, we consistently follow the input-based approach, defining augmentation (as well as automation) based on the roles that humans and AI play in performing a given task, regardless of the performance implications.[6]

A second source of ambiguity is the conflation of different levels of analysis. For example, if AI automates a task A within a worker's task portfolio, the worker is displaced from that specific task but may become more efficient in a complementary task B. While this may increase the worker's overall productivity, labeling this effect as "augmentation" conflates

[6] This input-based approach follows the original conceptual logic of augmentation as pioneered in scholarship on human-computer systems (Engelbart, 1962). In this classical sense, technology is viewed as a cognitive extension that increases the capability of an individual to approach a task.

task-level automation with its downstream effects on other tasks and the broader job (Raisch & Krakowski, 2021). Similarly, the notion that "one person's automation is another's augmentation" (Agrawal et al., 2023) is not paradoxical when we recognize that it typically refers to persons working in different occupations (e.g., a manager and her assistant). In this paper, we define automation and augmentation strictly at the level of a given task. This enables us to theorize the links between task characteristics and AI deployment at a consistent level of analysis (Puranam, 2018) and to abstract from dynamics that may play out across tasks, individuals, and occupations (Acemoglu & Restrepo, 2019; Raisch & Krakowski, 2021).[7]

In the next section, we refine our definition of automation and augmentation as distinct human-AI architectures for a given task, reflecting organizations' choices regarding how to allocate task instances to humans and/or AI.

2.2 Automation and augmentation: Independent versus joint human-AI architectures

Tasks typically entail multiple *task instances*, which refer to the performance of the task on a particular object, at a particular point in time, or in a particular place (Puranam, 2018). The customer service task of "answering phone calls", for example, is composed of answering many individual phone calls, each of which is a task instance (Brynjolfsson et al., 2023; Jia et al., 2023). The financial task of "making stock recommendations" consists of analyzing many different stocks and deciding on the recommendation for each one (Puranam, 2021). The research task of "collecting data on birds" consists of recording bird observations at different locations and at different points in time (Koehler & Sauermann, 2024). Organizations decide

[7] Although such temporal and aggregate effects are beyond the scope of our paper, they will, of course, have important implications for the macro-level labor market impacts of AI (Acemoglu et al., 2025).

how to allocate these task instances to humans and/or AI for execution, including whether humans and AI work on the same instances jointly or on different instances independently (Puranam, 2021). This allocation of task instances is captured in three possible "human-AI architectures" for the task: Human-only, automation, and augmentation (see Table 1).

Human-only is the baseline architecture where all task instances are performed solely by humans. In this case, the organization relies entirely on human cognition and material agency to complete the entire set of task instances and no AI is involved. For example, a bank may ask its human customer service agents to answer a series of phone calls without any involvement of AI, or an astronomer may ask her PhD student to classify a set of images of galaxies without any involvement of AI.

When using *automation*, the organization allocates task instances to AI for independent execution. In the extreme case of full automation, the entire set of instances is performed autonomously by the machine, e.g., if the organization uses an AI that has been fully pre-trained outside of the organization. In other cases, an organization may use partial automation – allocating some task instances to humans and others to AI for independent execution. Astronomers in the project *Galaxy Zoo*, for example, asked humans to classify a subset of images of galaxies to generate training data that enabled the AI to automatically classify all remaining task instances (Trouille et al., 2019). Regardless of whether automation is full or partial, we will classify cases where AI performs task instances independently as automation.[8]

When using *augmentation*, the organization designs a joint human+AI workflow: Humans and AI work on the same task instances, bundling their respective capabilities and inputs to perform an instance. For example, a hospital may design the workflow such that AI provides

[8] Of course, AI pre-trained on external data still required the human execution of task instances outside the organization. As such, the notion of full automation of a task is only meaningful when defined within the boundaries of a given organization.

preliminary diagnostic labels for MRI scans which a human then verifies and contextualizes (Lebovitz et al., 2022). Similarly, an organization may ask AI to generate real-time suggestions for how customer service agents should respond in a particular phone call (Brynjolfsson et al., 2023). Our definition of augmentation also applies to cases where humans and AI produce independent estimates or judgments for the same task instance, which the organization then aggregates to arrive at a final prediction or decision for that particular instance (Puranam, 2021). Again, augmentation may be full or partial, depending on whether the joint human-AI workflow applies to the full set of task instances or only to a subset. Regardless of whether augmentation is full or partial, we will classify cases where AI performs task instances jointly with humans as augmentation.[9]

Note that partial automation (humans perform some task instances, AI performs others) and augmentation (humans and AI perform the same task instances) both involve humans and AI in the overall task. However, the critical distinction is whether humans and AI work on different task instances independently (e.g., image classification in *Galaxy Zoo*) or whether they work on the same instances jointly (e.g., AI recommendations for doctors). This matters because the joint execution of an instance allows humans and AI to integrate their respective skills or judgments to complete that instance, while independent execution does not entail such an integration (Puranam, 2021).

--- Table 1 here ---

[9] Given that automation and augmentation can be partial, they are not mutually exclusive. For example, an organization could ask AI to classify all images in a set and retire those images that AI classifies with high confidence. Images classified with low confidence could be passed on to humans, along with the algorithmic classification as a recommendation. For simplicity, our conceptual discussion will assume that organizations choose either automation or augmentation (or human-only). Our empirical analysis allows for combinations of automation and augmentation, but such combinations are extremely rare.

We assume for now that the choice of human-AI architectures is driven by organizations' goal to minimize costs for a given level of quality (Acemoglu & Autor, 2011; Puranam, 2018). These costs include primarily the costs of developing and operating AI technology as well as the costs associated with using human labor (e.g., recruiting, wages, managerial oversight). Automation will reduce labor costs but whether it is more efficient than augmentation or a human-only architecture depends on the costs of the AI technology relative to those of the labor it replaces. Augmentation becomes the efficient choice when the AI's cognitive support increases human speed or performance enough to reduce labor cost by more than the additional costs of AI. In the next section, we leverage the task-based model (Acemoglu & Autor, 2011; Autor et al., 2003) to theorize how these design trade-offs depend on task characteristics, where the mapping between task requirements and AI capabilities determines the underlying cost structure of alternative architectures.

2.3 Task characteristics as predictors of human-AI architectures

We focus on two task characteristics that have long been central to the task-based model: task routinization and the physicality of work. The mapping between these task characteristics and current AI capabilities will determine the feasibility and implied costs of deploying AI for automation or augmentation (Autor et al., 2003; Brynjolfsson & Mitchell, 2017).

2.3.1 Routinization

Routinized tasks involve the execution of repetitive operations in a largely stable environment, where the same stimulus requires the same response over time. Such tasks can typically be codified, enabling organizations to provide explicit instructions for machines (Autor, 2015; Pentland & Rueter, 1994). In contrast, mastery of non-routine tasks often relies on tacit knowledge that is acquired through experience but cannot be fully expressed in

symbols or rules – a limitation known as Polanyi's Paradox (Autor, 2024; Autor, 2015). Consequently, the task-based model predicts that prior technologies such as robots and computers are more likely to automate routinized tasks than non-routinized tasks because these technologies typically require explicit instructions and face difficulties adjusting to changing environments.

It is not obvious that the link between routinization and automation also applies in the context of AI. Indeed, many observers claim that routinization and codification are much less important because algorithms can replicate human performance without receiving explicit instructions (Aghion et al., 2018; Brynjolfsson & Mitchell, 2017; Tschang & Almirall, 2021). For example, an AI can identify patterns in data on past human decisions and then make similar decisions without explicit instructions on how to evaluate decision options. We argue, however, that it is still easier for AI to perform routinized vs. non-routinized tasks, even though the reasons are likely different than for prior technologies. Among others, routinized tasks tend to have clearer performance criteria that can be targeted in training, relevant variables are more likely to be measured, training data tend to be larger, and there will be less model drift due to a changing environment (Choudhury et al., 2020; Jia et al., 2023; Raisch & Fomina, 2023; Sahiner et al., 2023).

Thus, we suggest that even though AI may be able to perform less routinized tasks in principle, its performance and cost advantages over humans will be greater in routinized tasks for reasons such as greater data availability or less model drift (Choudhary et al., 2025). Thus, we expect that task routinization will shape organizations' decisions to automate also in the context of AI. For non-routine tasks, the costs of building, monitoring, and updating an AI model that delivers sufficient quality will typically exceed the potential savings in labor cost such that a human-only workflow remains more efficient. AI-based automation becomes more viable as routinization increases. Organizations should be most likely to enjoy "cheap

prediction" (Agrawal et al., 2018) and assign task instances for independent execution to AI when tasks are highly routinized.

Hypothesis 1: *Tasks with high routinization are more likely to be automated than tasks with low or medium levels of routinization.*

Conversely, augmentation is most likely to emerge as an organizational design at medium levels of routinization. In such tasks, AI may offer valuable cognitive inputs that reduce humans' cognitive load or increase performance when used jointly with humans, offering advantages over a human-only design. Given that AI is less reliable than for task with high routinization, however, the organization is more likely to bundle AI inputs with human judgment than to automate. For example, radiologists may receive algorithmic recommendations based on large amounts of training data but incorporate in their final diagnoses additional tacit knowledge as well as ethical or legal considerations that may be more unique for each case (Lebovitz et al., 2022). Similarly, patent examiners may receive algorithmic recommendations for relevant prior work but can use their domain-specific expertise to account for changes in language and applicants' strategic invention of new multiwords (Choudhury et al., 2020).

Assuming that automation promises greater efficiency gains than augmentation due to lower labor costs or higher speed, organizations should automate when tasks are sufficiently routinized to enable reliable automation and should augment if tasks are not sufficiently routinized to automate but still routinized enough for AI to support humans (Huang & Rust, 2018). In contrast, low-routinization tasks will typically present significant barriers for AI, likely rendering augmentation inefficient and leading organizations to choose a human-only architecture.

Hypothesis 2: *Tasks with medium levels of routinization are more likely to be augmented than tasks with low or high levels of routinization.*

We note that these predictions are contingent on the current state of technological maturity. They are premised on the assumption that – despite rapid improvements in AI's ability to learn from data rather than relying on explicit rules – the lack of routinization remains a relevant constraint, albeit for different reasons than before. While prior technologies were limited by humans' difficulty of articulating tacit knowledge (Polanyi's Paradox), AI is limited by insufficient data and data shifts that impede reliable independent prediction especially in non-routinized tasks (Choudhary et al., 2025). Of course, the boundary of what can be automated will move as AI evolves, likely reducing the role of task routinization. As such, our test of Hypotheses 1 and 2 will show whether routinization remains a structural constraint for current AI, while the strength of this relationship offers a gauge for how far we may be from the point where routinization ceases to matter.

2.3.2 Physicality

While the task-based model and studies on AI exposure explicitly incorporate physical requirements of work (Autor et al., 2003; Brynjolfsson & Mitchell, 2017; Felten et al., 2021), this dimension has been largely overlooked in recent organizational literature on the role of AI. This may reflect that research in this area has primarily focused on knowledge work and decision-making (Dell'Acqua et al., 2023; Puranam, 2021; Raisch & Fomina, 2023), abstracting away the sensorimotor requirements that remain central to many real-world tasks such as transporting objects, visiting customers, collecting environmental data in nature, or performing surgery (Autor, 2015; Beane, 2019). We argue that attention to task physicality is important because it will shape the feasibility and costs of different human-AI architectures.

The link between task physicality and human-AI architectures is grounded in an inherent technical limitation of software-based AI: It lacks the ability to independently move through material space and to manipulate physical objects (Agrawal et al., 2018; Felten et al., 2021).

As such, the alignment between task requirements and AI capabilities is expected to be highest when tasks are cognitive and can be completed within digital or symbolic environments, and such tasks should exhibit the highest levels of automation.

By contrast, any task involving a physical component presents a structural hurdle for software-based AI. For purely physical tasks, requiring primarily sensorimotor skills but little information processing, the implementation of AI is neither feasible nor desirable: there is little cognitive value-add of AI, while the technical costs of bridging the digital-material gap would be prohibitive. Consequently, organizations will continue to organize physical tasks using human-only architectures.

Hypothesis 3: *Tasks that are primarily cognitive are more likely to be automated than hybrid or primarily physical tasks.*

Theoretically most interesting are hybrid tasks. In these tasks – such as collecting data on birds while navigating a forest – information processing and material agency are closely intertwined in performing a task instance.[10] It is obvious that such tasks cannot be automated by a software-based AI, but it is less obvious why organizations would not just rely on the baseline human-only architecture, given that humans possess both the cognitive and sensorimotor skills required. We argue that joint performance of the same task instances by humans and AI can be a superior design solution because it enables the bundling of diverse inputs to increase performance or efficiency. More specifically, while humans can provide essential material agency and may possess important tacit understanding of the (physical) context, they are often bounded with respect to their access to and processing of large volumes of codified information. AI, in contrast, will often have superior access to codified information

[10] Although rarely considered explicitly in management or economics, hybrid or "dual" tasks are well-researched in physiology and medicine (Lauenroth et al., 2016).

as well as the ability to process such information to make relevant predictions (Choudhary et al., 2025; Puranam, 2021). Bundling these diverse strengths by using a joint human+AI architecture may be beneficial especially in hybrid tasks because such tasks tend to require a more diverse set of inputs than cognitive or physical tasks. When collecting data on birds in nature, for example, a human's sensorimotor skills are required to navigate to physical locations and capture visual data of a bird, but AI could suggest a taxonomic identification of the observed bird based on a large amount of training data. This identification could be accepted or modified by the human based on her understanding of the particular local context or situational factors that are not accessible to AI (Kelling et al., 2013).

Of course, AI may also augment rather than replace humans in purely cognitive tasks, especially if such tasks involve activities that current AI still struggles with (Choudhury et al., 2020; Kittur et al., 2019). Given the strong information processing capabilities of AI, however, we expect that augmentation is more likely in hybrid than in cognitive tasks because humans and their different kinds of capabilities and knowledge remain more critical in the former than the latter. Interestingly, while rapid improvements in the cognitive capabilities of AI will increase the share of cognitive tasks that are amenable to automation, such improvements are unlikely to alleviate the constraint of physicality such that the augmentation of hybrid tasks may be a stable design choice rather than a transitory state towards full AI-based automation.

Hypothesis 4: *Hybrid tasks (requiring both cognitive and sensorimotor inputs) are more likely to be augmented than primarily cognitive or primarily physical tasks.*

3 Empirical context

3.1 Crowd science

We test our predictions in the context of crowd science (CS): Research projects that recruit participants through an open call for contributions (Franzoni & Sauermann, 2014; Haklay et

al., 2021; Lyons & Zhang, 2019). CS projects have their own structures and processes and can involve thousands of participants working towards project-level goals, making them meaningful organizations (Puranam, 2018). Projects are typically led by professional scientists, including at top tier institutions such as Cornell, Oxford, and Stanford University (Andreasson et al., 2022; Khatib et al., 2011; Trouille et al., 2019). Crowd members are not paid and may include other scientists but are primarily individuals without formal scientific training who contribute their time, effort, and relevant knowledge.

Crowd science projects are a suitable research context for several reasons. First, they typically involve participants in one well-defined primary task rather than broader bundles of tasks (as do most jobs). For example, the main task for participants in the project *eBird* is to observe birds in nature and submit records of the species and their numbers. In *Galaxy Zoo*, crowd members classify images of galaxies with respect to properties such as their shape, brightness, and irregular features. In *Eterna*, crowd members design novel RNA structures with desirable properties. This allows us to measure both task characteristics and human-AI architectures at the level of clearly defined tasks across many organizations. As such, we complement prior work that has studied AI workflows within single organizations or for single tasks (Dell'Acqua et al., 2023; Jia et al., 2023; Lebovitz et al., 2022), as well as studies that measure AI deployment using firm-level proxies such investments in IT, survey-reported AI use, or job postings that include AI-related skills (Babina et al., 2024; McElheran et al., 2024).

Second, projects exhibit significant heterogeneity with respect to our focal task characteristics. This variation stems from the diversity of scientific activities required in different fields and for different research questions – ranging from purely cognitive data processing, such as classifying astronomical images or folding protein structures, to physical fieldwork, such as collecting soil samples in varied terrains. The task design is typically defined by organizers as part of the planning process (Poetz & Sauermann, 2024; Trouille et al., 2019),

but some projects also allow participants to shape organizational decisions through mechanisms such as a player council in *Eterna* or online discussion boards. The heterogeneity in task characteristics allows us to exploit cross-project variation to test our predictions.

Third, organizers often find it difficult to recruit and retain the volume of participants required to achieve their research goals (Lyons & Zhang, 2019; Sauermann & Franzoni, 2015). Projects also incur substantial costs of training, onboarding, and managing a dispersed workforce (Balázs et al., 2021; Koehler & Sauermann, 2024). Even though participants do not get paid, labor is not free and crowd science organizers have an incentive to use it efficiently. This makes choices regarding AI deployment an important decision for them (Ceccaroni et al., 2019; Trouille et al., 2019), enabling us to study how organizations design human-AI architectures depending on the nature of the task that needs to be accomplished.

Finally, crowd science projects tend to be very open and transparent (Franzoni et al., 2022). Most projects have websites that serve to recruit participants and can also be used to collaborate on problems or submit contributions. Some projects discuss their organizational setup in the methods sections of scientific articles, while others publish dedicated reports detailing how they work (e.g., Trouille et al., 2019; Zevin et al., 2024). This high degree of openness allows us to observe task characteristics and AI use in a large number of projects, complementing prior work that has studied AI adoption within a single organization or for a single task (e.g., Brynjolfsson et al., 2023; Choudhury et al., 2020; Jia et al., 2023).

3.2 Learning as an additional organizational goal

An interesting feature of crowd science is that some projects pursue not only scientific outputs (e.g., data, publications, new protein designs) but also a secondary objective – the learning and education of participants. This may include learning about how to perform specific tasks such as taxonomical classifications, learning about more general scientific methods and

processes, as well as learning and awareness of certain natural phenomena or environmental challenges. The *Rescue A Reef* project, for example, allows crowd members to participate in research expeditions to "learn about the critical role that coral reefs play in our environment and economy, threats they face, and tools available for their conservation and recovery."[11] The project *DeVOTE* involves citizens in research on the meaning of voting and states that "In this way, exchange between the interested public and academia is facilitated and academia becomes tangible through citizen science".[12] *Waterwatch Victoria* involves crowds in environmental monitoring, in the process offering "school and community education and awareness-raising activities".[13] Although learning objectives may be more salient in crowd science than in other contexts, learning has been problematized in several recent discussions of AI. In particular, there is a concern that the automation of tasks – especially routinized tasks often performed by entry-level workers – may deprive people of opportunities to practice their craft, undermining the development of human capital (Beane, 2019; Ellis & Bindley, 2025; Garicano & Rayo, 2025).

We argue that organizations that value learning may be less inclined to delegate task instances independently to AI because automation replaces human labor, eliminating opportunities for human learning-by-doing. In contrast, augmentation may support learning goals through two complementary pathways. First, by keeping the human in the loop to perform the task instance jointly with the AI, augmentation ensures continued engagement and learning-by-doing (He et al., 2024). Second, a joint workflow where the AI provides recommendations, probabilities, or explanations can enable the human to learn from the AI (Choudhary et al.,

[11] https://rescueareef.earth.miami.edu/

[12] https://www.votemeanings.eu/participate-now/activity-overview

[13] https://scistarter.org/waterwatch-victoria

2025). For example, contributors recording data in nature can enhance their taxonomic skills by engaging with AI's recommended classifications for species they have not seen before, or by using algorithmic guidance to identify the features most relevant to differentiate species that are easy to confuse (Koehler & Sauermann, 2024).

We note that attention to learning goals is not inconsistent with the task-based model, which includes a flexible objective function that is often assumed to focus on task efficiency but that can also incorporate other types of goals (Acemoglu & Restrepo, 2020). Although not explicitly theorized, we therefore explore empirically whether projects that also pursue learning goals make different decisions regarding human-AI architectures than projects that focus solely on scientific output. We will discuss the generalizability of our results beyond the context of crowd science in section 6.2.

4 Data collection and measures

4.1 Data collection

To identify a large and diverse set of projects, we collected information starting with two sources. First, we collaborated with SciStarter.org, the largest public catalog of crowd science projects in terms of the number of projects listed. SciStarter has a focus on the USA but also covers projects from other English-speaking countries. These projects span a range of fields, with an emphasis on ecology, biology, climate science, and related areas. Project profiles listed on SciStarter provide information about the purpose of the research, the task to be performed, as well as a link to the main project website. We received from SciStarter a list with all 1,504 projects shown as active in the Spring of 2023.

Second, we identified additional projects on Zooniverse.org, the largest crowd science platform in terms of the number of registered users (2.7 million as of June 2024). This platform hosts projects in a broad range of fields including astronomy, biology, medicine, physics, and

the humanities. The primary tasks for participants involve data processing, typically by classifying images or videos with respect to scientifically relevant characteristics. Zooniverse had 108 active projects at the time of our data collection, of which 35 were also listed on SciStarter.org. Combining the unique projects from the two sources, our starting sample included 1,577 projects.

One co-author and a research assistant ("coders") collected detailed information on these projects using different sources such as profiles on SciStarter.org, projects' main websites, blogs and reports, as well as research articles published by the projects and identified on project websites or through Google Scholar. In a few cases, the coders also signed up for projects as participants to better understand tasks and the use of AI. Both coders investigated the full set of projects independently, using detailed guidelines that were developed and pre-tested by the authors (see below for details on individual variables). The coders met via Zoom repeatedly to combine their evidence and agree on the coding of human-AI architectures and other key variables. Another co-author helped resolve unclear cases throughout the coding process.[14]

Upon detailed examination by the coders, some of the projects were deemed ineligible, including 263 projects that turned out to be inactive at the time of the data collection. We dropped an additional 109 projects because they were not primarily about scientific research or did not actively involve crowd members in tasks. Examples include projects focused on science education or policy advocacy rather than research (e.g., *NOVA Labs* projects), projects where participants are not active researchers but research subjects that complete surveys on themselves or donate personal specimens (e.g., *Personal Genome Project*), and projects where

[14] The coders had an average agreement of more than 80% in their initial coding of key variables. The coders' primary task was evidence gathering, i.e., to find relevant information on the use of AI or on the tasks performed by participants. As such, the initial codings were not averaged but the underlying evidence was combined to determine the most appropriate coding.

participants do not perform work but only contribute other resources such as computing power (e.g., *LHC@home*). Our final sample consists of 1,205 projects.

4.2 Measures

Table 2 shows summary statistics and Appendix Table A3 shows correlations.

4.2.1 Independent variables: Task characteristics and learning goals

For each individual project, the coders carefully read project descriptions and other sources to identify the primary task. These tasks are at a level of aggregation that is meaningful to both organizers and participants, and that permits reliable measurement (e.g., reporting roadkill in *Save the Tasmanian Devil*, identifying distinct features of letters in different scripts in *Glyph*, classifying galaxies in *Galaxy Zoo*). Although some projects also include secondary tasks such as helping write a paper, these secondary tasks are typically performed by a small subset of participants (Wellington-Oguri et al., 2020). Our measures of task characteristics (and human-AI architectures) focus on the primary tasks only.

Task routinization. The coders rated task routinization on a 3-point scale (1=low, 2=medium, 3=high). In doing so, they considered the frequency and type of actions carried out by individual participants and the context in which these actions occur (Autor et al., 2003). Routinization is coded as high for very repetitive tasks performed in a largely constant environment, such as classifying images online in *Snapshot Safari*, identifying clogged blood vessels in mouse brains in *Stall Catchers,* or monitoring seismic activity repeatedly in the same location in *Quake Catcher Network*. Routinization is coded as low for tasks that are not repetitive or performed under very different conditions, such as performing statistical analyses in *Exoplanet Watch*, excavating and processing mammoth bones in *Mammoth Site Ice Age Explorers*, or folding complex proteins in *Foldit*. Routinization is coded as medium for tasks that are somewhat repetitive and may be performed under similar conditions. Examples include

making repeated observations in the *Illinois Butterfly Monitoring Network,* monitoring water quality in *Bellingen Riverwatch*, or evaluating videos of criminal trial evidence in *Justice BRD*.

Task physicality. We classified tasks based on the medium of execution and the requirement for material agency. Tasks were categorized as primarily cognitive if they are performed entirely in an online environment, where the focal activity is information processing. Examples include molecular design in *Eterna*, spatial tracing of fishing activity in *Global Fishing Watch,* and morphological classification of galaxies in *Galaxy Zoo*. Conversely, tasks were coded as primarily physical when performed offline, requiring material agency without significant digital mediation, such as the collection of soil samples, microplastics, or insects for later analysis (e.g., *LA Spider Survey*). Tasks were classified as hybrid if they involve both offline and online elements, requiring material agency and digital information processing within the same task instance, such as observing wildlife in the field and recording observations via online interfaces (e.g., *eBird*, *Monarch Larva Monitoring Project*). We recognize that the mapping between task physicality and the online vs. offline medium of task execution is not perfect, e.g., because offline tasks may also involve cognitive elements. For our purposes, however, it is sufficient to make the reasonable assumption that cognitive elements are more important than physical elements in online tasks, while physical elements are more important than cognitive elements in offline tasks.

Learning goals. To assess whether projects pursue not only scientific goals but also goals related to the learning and education of participants, the coders reviewed sources such as project websites, blogs, as well as reports and publications for explicit goal and mission statements. This variable is coded as 1 if projects explicitly mentioned goals related to learning, education, or building awareness among participants (see section 3.2 for examples). In many cases, such goals were also manifested in tangible project features such as courses, workshops, or online training materials. Of course, one may be concerned that verbal statements regarding

learning goals are simply marketing efforts by projects seeking to appear in a positive light. Although we cannot rule out this possibility, only a minority of projects state learning goals, suggesting that any potential marketing benefits are not widely expected. Moreover, if some projects only pretend to pursue learning goals, then the empirical relationship between this measure and AI deployment decisions should be attenuated, making any observed effects potentially more informative.

4.2.2 Dependent variables: Human-AI architectures

Following our conceptual discussion in section 2, we coded organizations' use of different human-AI architectures focusing on the roles that humans and AI play in performing task instances (input-based, regardless of productivity). Towards this end, the coders searched all available sources for evidence that a project uses AI in performing instances of the primary task. The coders first looked for relevant terms such as machine learning, artificial intelligence, deep learning, models, training data, and algorithm. If such terms appeared, the coders tried to find more detailed information to verify the use of AI and to learn about its concrete role in the workflow. In some cases, the coders participated in the project to better understand how AI is deployed, e.g., to see algorithmic recommendations given to participants in a biodiversity monitoring project.[15] Decisions about AI use in crowd science projects are typically made by project organizers and technical staff who develop relevant features and embed them into workflows (Fortson et al., 2024; Parrish et al., 2019). Although individual participants in some projects may have the choice not to use available features, we focus on organizers' decisions to deploy AI at the project level rather than on individual users.

[15] One of the coders co-founded an AI startup and drew on this experience in identifying relevant AI applications.

Automation. Applying our definition summarized in Table 1, the coders assessed whether a project allocated some or all instances of the primary task to AI for independent execution. *Galaxy Zoo*, for example, uses AI to independently classify images of galaxies, using different task instances classified by human participants as training data (Trouille et al., 2019). In *Gravity Spy*, an AI pre-classifies all images and the subset that is classified with a low level of confidence is passed on to experienced crowd members for human rather than AI processing (Zevin et al., 2024). In *Quantum Minds*, participants solve problems related to atomic state transition optimization, and the lead scientists then analyze these solutions to develop algorithms that can independently solve new problems of the same type. Note that we could not track the evolution of projects to see whether and how AI has taken over task instances over time. Rather, we rely on published information about how AI has been used (e.g., Zevin et al., 2024) as well as on clearly stated intentions. For example, *Quantum Minds* states that "Getting a deeper understanding of the human problem solving will help scientists create better algorithms for solving quantum problems",[16] while *Industrial Smoke Hunting* explains: "Our goal is to train an Artificial Intelligence system to detect industrial smoke pollution automatically. To do this, we need a vast amount of data, which will require a considerable human effort."[17] In our sample, 7.72% of projects use AI for automation (dummy = 1).

Augmentation. Following our definition summarized in Table 1, the coders assessed whether AI and humans work on the same instances of the primary task, integrating inputs from both humans and AI using a joint workflow. In *eBird*, for example, an AI tool helps participants identify a given bird faster and with greater accuracy by suggesting possible classifications and by asking for additional verification if participants report a species that is unlikely to occur in

[16] https://www.scienceathome.org/games/quantum-minds/about/

[17] https://scistarter.org/industrial-smoke-hunting

a particular location or time of the year (Figure 1). In *Aurorasaurus*, AI helps crowd members make observations of northern lights by recommending the best time to observe an aurora in a particular location. The project *Phylo* solves multiple sequence alignment problems by asking humans to improve AI-generated solutions. Thus, the dummy variable A*ugmentation* captures the concrete workflow architecture designed by project organizers, taking the value of 1 for projects where human and AI inputs are integrated to perform individual task instances and produce single units of output (e.g., a particular image classification, a particular sequence alignment). In our sample, 8.80% of projects use AI for augmentation.[18]

--- Figure 1 about here ---

We recognize three important limitations of our measurement of AI. First, the coders had to rely on public information from sources such as websites, reports, and academic publications. Given that they coded AI use only if there was sufficient evidence, it is likely that we missed cases of AI. Second, all crowd science projects involve humans in at least some task instances and there are no projects where the primary task is fully automated, preventing us from studying this extreme end of the spectrum. Third, we do not have sufficient information to reliably measure additional nuances such as the intensity of AI use (e.g., the share of task instances that are performed by AI in projects using automation; the relative contribution of AI and humans in joint human+AI architectures) or the exact sequencing of human and AI activities. Despite the lack of such nuance, our measures improve significantly upon aggregate proxies of AI adoption that cannot distinguish between automation and augmentation (Babina et al., 2024;

[18] Only four projects use AI to both automate and augment (e.g., some task instances are performed independently by AI, others jointly by humans and AI). The number of such projects is too small to analyze them separately, and values of 0 in the dummies *automation* and *augmentation* can essentially be interpreted as the use of a human-only architecture as the baseline.

McElheran et al., 2024), and our focus on whether task instances are performed jointly or independently matches exactly our definition of different human-AI architectures and our theorizing regarding their likely use (Section 2). Of course, future research should explore alternative approaches that are able to address the limitations of our measures.

4.2.3 Control variables

Platform affiliation. Although many projects are independent, others reside on crowd science platforms such as Zooniverse, Biocollect, iNaturalist, or Science at Home. Platforms typically provide access to an established user base and may also offer shared technical infrastructure such as templates to implement common tasks (e.g., classify images). Most relevant in our context, platforms may provide access to shared AI tools, reducing projects' costs of implementing AI.[19] We coded platform membership based on information provided on project websites and by searching for indirect evidence such as shared domain names or shared member logins (dummy = 1 if a project resides on a platform).

Project age. The use of AI may depend on project age, e.g., because older projects have accumulated more data that AI can learn from or because younger projects may find it easier to integrate new technologies into their workflows. For projects sourced from SciStarter, we received age information directly from SciStarter (years since the first listing of the project). For Zooniverse projects, we coded age manually (years since project start). We use six categories: 0-1 year, 2, 3, 4, 5, 6+ years.

Secondary tasks. Some projects also engage participants in secondary tasks such as donating funding or helping analyze crowd-generated data. Such tasks are typically performed by only

[19] We code AI deployment independently from platform membership. The exceptions are projects on iNaturalist, which by design all use the same AI tool to support contributors with species identification (coded as using augmentation).

few participants and may not always be visible to external observers. Although such tasks are not considered in our coding of task characteristics or human-AI architectures, we include a dummy variable that takes the value of 1 if we found evidence of any secondary tasks.

Geographic scope. Projects differ in their geographic scope. Some projects, for example, study air pollution or biodiversity change in a particular city or region. Others address more general research questions, collecting data and involving crowd members at a national or international scale. Geographic scope may affect recruiting efforts and project size but may also capture the diversity of data (e.g., the global project *eBird* deals with a much broader range of species than a local bird observation project in the state of North Carolina), which may have implications for the benefits and costs of using AI. We received information on geographic scope directly from SciStarter and we coded it manually for Zooniverse projects (local/regional projects versus national/international projects). We created a separate dummy variable for projects where geographic scope could not be determined reliably.

Research field. There may be systematic field differences in the benefits and costs of AI deployment, access to necessary resources, as well as the availability of human contributors. To control for such differences, we coded research fields based on project profiles on SciStarter or Zooniverse. The variable *Mainfield* captures the primary field. The variable *Secondfield* captures any relevant secondary fields (e.g., for projects spanning multiple fields, interdisciplinary projects), with a separate category for projects without a secondary field. Table 2 shows that a large share of projects are in biology as well as the earth and environmental sciences, which reflects that these fields are at the frontier of using crowd science approaches (Hecker et al., 2018; National Academies, 2018). Unfortunately, there is no data on the universe of crowd science projects, preventing us from studying sample representativeness formally.

--- Table 2 about here ---

5 Results

5.1 Mean comparisons

Figure 2 visualizes the share of projects using augmentation and automation architectures by the degree of task routinization (Panel 1) and task physicality (Panel 2). Panel 3 adds exploratory evidence regarding differences depending on the presence of learning goals. Regressions below will test the emerging differences statistically while accounting for controls and correlations between independent variables.

Panel 1 shows that automation is most likely in highly routinized tasks, consistent with Hypothesis 1. Interestingly, some projects also use AI to automate tasks with low routinization. A closer look at these cases shows that several involve crowd members in experimentation or problem solving and then analyze human strategies to develop better algorithms. In other words, even when problem solving is not routinized and codified, AI can look across many individuals and learn from their collective problem-solving behaviors (Aghion et al., 2018; Tschang & Almirall, 2021). Once algorithms have improved sufficiently, they can take over remaining task instances.[20]

Augmentation is most likely in tasks with medium routinization, consistent with Hypothesis 2. An example is *Frog ID*, which provides participants with AI support in the task of recording frog observations. Although some projects also employ joint human+AI architectures in high routinization tasks (e.g., an AI in *Living with Machines* highlights potentially relevant words

[20] The project *Foldit* is a particularly successful example: https://fold.it/portal/node/2008706.

in old newspaper articles to facilitate transcription), this is less common, perhaps because automation is often a better option for such tasks.[21]

Panel 2 shows that automation is most likely in cognitive tasks, consistent with Hypothesis 3. Indeed, our sample includes no projects where AI is used to automate physical or hybrid tasks – consistent with the premise that AI is unable to independently perform tasks that require material agency (Agrawal et al., 2018). Augmentation is most likely in hybrid tasks, supporting our conjecture that such tasks tend to require more diverse inputs than physical or cognitive tasks and can, therefore, benefit the most from bundling the capabilities of humans and AI (Hypothesis 4). A prominent example is *eBird*, which asks participants to make observations in different physical locations but facilitates the classification and recording of these observations using an AI tool.

Panel 3 shows exploratory evidence that automation is less likely in projects with learning goals than in projects without learning goals. For example, the *Mystic River Herring* project does not automate the routinized task of counting herrings in videos even though automated solutions exist (Luo et al., 2015). This partly reflects that this project also pursues educational goals; automating the task would undermine the project's ability to achieve its learning goals by engaging participants actively in research.[22] In contrast, Panel 3 shows no difference with respect to the use of AI for augmentation, suggesting that there is no inherent conflict between learning goals and the use of joint human+AI architectures.

--- Figure 2 about here ---

[21] *Living with Machines* explains: "We need your help because computers can't match human abilities for understanding historical texts." See https://www.zooniverse.org/projects/bldigital/living-with-machines/about/research.

[22] See https://www.mysticherring.org/. Among others, this project partners with schools, stating that "Students will perform herring counts, develop critical thinking skills, improve environmental literacy, and participate in stewardship opportunities."

5.2 Main regressions: Task characteristics and human-AI architectures

We estimate regressions with *Automation* and *Augmentation* as dependent variables, using linear probability models with robust standard errors (Angrist & Pischke, 2008).

Models 1-4 in Table 3 focus on task characteristics individually, showing patterns that are consistent with Figure 2. However, task routinization and task physicality are correlated: The cognitive tasks that are common in our sample tend to be highly routinized (e.g., image classification, transcription), while physical and hybrid tasks such as raising caterpillars or measuring water quality in different locations are less routinized. To account for this correlation, models 5 and 6 include both task characteristics jointly. Although the magnitude of coefficients is reduced, model 5 continues to strongly support H1 (highest automation in highly routinized tasks) and H3 (highest automation in cognitive tasks). Model 6 provides partial support for H2 (highest augmentation in medium-routinization tasks) in that augmentation remains significantly more likely in medium than in high routinization tasks. However, the difference between low and medium routinization is not statistically significant. Similarly, support for H4 (highest augmentation in hybrid tasks) is partial in that augmentation remains significantly more likely in hybrid than physical tasks, while the difference between hybrid and cognitive tasks is not significant.

In model 7, we probe the importance of distinguishing between automation and augmentation by collapsing the two measures into a single measure *AnyAI* (dummy coded =1 if a project uses AI in any role). This measure can be interpreted as a proxy for the (realized) more general AI exposure of different types of tasks (see Eloundou et al., 2024; Felten et al., 2021; Sytsma & Sousa, 2023). We find that routinization does not predict *AnyAI*, even though it predicted both augmentation and automation in the main regressions. The reason is that high levels of routinization predict higher levels of automation but lower levels of augmentation

(models 1 and 2). Averaging these contrasting relationships by using the aggregate measure obscures important differences. Similarly, the coefficients of task physicality in regressions of *AnyAI* (model 7) give misleading impressions of the strength and significance of the relationships observed in separate regressions of *Automation* and *Augmentation* (models 3 and 4).

--- Table 3 about here ---

5.3 Exploratory analysis: The potential role of learning goals

Table 4 shows regressions that explore the relationship between human-AI architectures and learning goals. Reinforcing the mean comparison in Figure 2, model 1 shows that projects with learning goals are significantly less likely to automate by assigning task instances to AI for independent execution. In contrast, Model 2 shows no difference with respect to the use of augmentation (joint human+AI architectures).

Models 3 and 4 include learning goals and task characteristics simultaneously. The coefficients of task characteristics are virtually unchanged compared to the main results in Table 3. The negative coefficient of learning goals in the regression of automation, however, becomes insignificant, likely reflecting that learning goals and task characteristics are correlated. To further explore this possibility, we use ordered logistic regressions to regress task characteristics on learning goals. These regressions confirm that tasks in projects with learning goals tend to be less routinized (model 5) and less likely to be purely cognitive (model 6). The interpretation is not clear. If we assume that task characteristics are fully determined by technical requirements of the project, then it could be that projects requiring certain types of tasks see different opportunities to accomplish learning goals. More realistically, however, task characteristics are not exogenous and partly reflect organizers' goals (Puranam, 2018).

Organizers who seek to educate participants and build human capital, for example, may design tasks that are less routinized and more varied with respect to their physical and cognitive nature (see Hackman & Oldham, 1976; National Academies, 2018). Following this perspective, task characteristics may mediate an effect of learning goals on automation. Future research is required to investigate the causal nature of these relationships in more detail.

--- Table 4 about here ---

5.4 Robustness check: Controlling for project size

Finally, we explore the robustness of our results to controlling for project size, proxied by the number of participants. The role of size is not clear ex ante. On the one hand, size may affect the design of human-AI architectures if larger projects have a larger volume of work to be performed or command greater resources and expertise to develop AI technologies. On the other hand, AI deployment may also affect project size, e.g., if automation reduces the required number of crowd members or if augmentation lowers skill requirements and enables broader participation (Brynjolfsson et al., 2023; Dell'Acqua et al., 2023; Koehler & Sauermann, 2024).

Extensive search for size information in sources such as webpages, blogs, and reports yielded data for 58.17% of the projects. We coded three size categories: small (<=1,000 participants; 50.36% of projects with size information), medium (1,001-10,000; 30.10%), and large (10,000+; 19.54%). We created an additional dummy variable equal to one if size information is missing.

Models 1, 3, and 5 reported in Table 5 show that project size has a positive relationship with automation, suggesting that positive effects of size on AI use (possibly via greater resources or volume of work) outweigh potential negative effects of automation on the number of participants. Size also has the expected positive relationship with augmentation (models 2, 4,

6). More importantly, the coefficients of task characteristics (and learning goals) are very similar to those in the main analyses, suggesting that our results are robust to potential confounding effects of project size.

--- Table 5 about here ---

5.5 Remaining potential for endogeneity

Before we turn to the general discussion, we note that our results are cross-sectional, yielding correlational but not causal evidence. In particular, we cannot identify to what extent task characteristics are determined by the intrinsic nature of research activities and to what extent they are shaped by organizers who design tasks differently for reasons such as resource constraints or the presence of learning goals (Hackman & Oldham, 1976; National Academies, 2018). Task characteristics would be important antecedents of AI deployment either way, but the second case would imply an even more fundamental role of organizational goals and constraints.

Relatedly, we control for several factors such as scientific field as well as project age and size, but we cannot rule out that other unobserved factors shape both AI adoption and our independent variables. For example, projects with greater technical expertise, institutional support, or access to funding may be more likely to design certain types of tasks and certain types of human-AI architectures, potentially confounding our results. Similarly, disciplinary norms and platform-level infrastructures may shape both task design and AI use in ways that are not fully captured by our field and platform controls. As such, we interpret our findings as descriptive patterns that are consistent with our theoretical framework rather than as causal estimates. Future research should complement our correlational evidence with a greater focus on identification, e.g., by tracking the evolution of projects over time, exploiting shocks such

as the arrival of new AI technologies or sudden cuts to university funding, or using experiments to study the effect of different task scenarios on organizers' design of human-AI architectures. We hope that our conceptual framework and initial empirical evidence using unique cross-project data will motivate and provide a useful starting point for such future work.

6 Discussion

6.1 Summary

Complementing prior work on the impact of artificial intelligence on workers and organizations, we developed and tested a conceptual framework to better understand when and how organizations deploy AI. Our framework bridges the task-based model from economics (Acemoglu & Autor, 2011) with the structural approach to organizational design (Puranam, 2018) to predict the use of different human-AI architectures, including automation (the allocation of task instances to AI for independent execution) and augmentation (the bundling of human and AI inputs to perform task instances).

Our data from more than 1,200 crowd science projects show that task routinization remains a key structural predictor of automation architectures. This suggests that even as modern AI moves toward advanced pattern recognition, the data richness and stability of routine environments still provide the clearest path for autonomous delegation. Conversely, organizations are most likely to design augmentation architectures when tasks have medium routinization, suggesting that AI input can meaningfully support humans in performing a given task instance – but current AI capabilities are insufficient for automation.

Task physicality is a powerful second predictor. Organizations are most likely to design automation architectures when tasks are primarily cognitive, where information processing occurs within digital or symbolic environments. In contrast, for hybrid tasks that require both cognitive processing and material agency, organizations tend to configure joint human+AI

architectures. This indicates that the requirement for sensorimotor skills in hybrid tasks remains a significant factor favoring the bundling of AI inputs with human labor.

Finally, our exploratory analysis suggests that AI deployment decisions may not only be shaped by task characteristics but also by organizational goals that go beyond task efficiency. In particular, we find that projects that value learning and human capital development among their participants are less likely to automate, while they are equally likely as other projects to use augmentation architectures that bundle human and AI effort.

6.2 Generalizability and boundary conditions

Crowd science has developed rapidly over the past decades, involving millions of participants, producing research published in top tier journals such as *Nature* and *PNAS*, and paving the way for a recent Nobel Prize. Although this makes crowd science novel and relevant, this context also warrants careful consideration of generalizability and boundary conditions.

A first consideration is that many CS projects involve small tasks of low complexity, which partly reflects that such tasks are suitable for parallel processing by a distributed workforce. Although our sample did not include large complex tasks such as running a clinical trial or writing a full research paper, traditional organizations also typically decompose such tasks before assigning them to individual workers or to AI (Brynjolfsson et al., 2023; Puranam, 2018). Thus, the tasks we study may not differ that much from the tasks typically performed by a significant share of the scientific and general workforce (Shibayama et al., 2015). Moreover, our sample shows considerable variation in task routinization and physicality, and including an even broader range of tasks should only strengthen the observed relationships between task characteristics and organizations' use of different human-AI architectures. Nevertheless, our study does not speak to the drivers of automation or augmentation when tasks are very complex, and future research on such tasks would be particularly valuable.

Second, crowd science participants are unpaid. This feature is important because access to low-cost labor may reduce organizations' incentives to implement AI, especially for the purpose of reducing labor inputs. Even without direct labor costs, however, crowd science projects have incentives to invest in AI for reasons such as reduced costs of attracting and managing participants, accelerating the completion of work, and higher quality when AI supports crowd members (Ponti & Seredko, 2022; Trouille et al., 2019). Although we believe that our key results generalize to contexts where labor is more costly, future research is needed on the cost side of AI adoption decisions, including the costs of involving humans, the costs of deploying AI, and how these costs differ when AI is used to automate versus augment (see Acemoglu & Restrepo, 2018; Svanberg et al., 2024).

Third, learning goals may be more prevalent in crowd science projects than in other organizations and it is not clear whether our exploratory results regarding the role of learning goals generalize to other contexts. At the minimum, recent discussions of the implications of automation for learning and human capital development among both scholars and business practitioners suggest that this question has more general relevance (Beane, 2019; Brynjolfsson et al., 2023; Ellis & Bindley, 2025; Garicano & Rayo, 2025). Future research is needed to study the salience of learning goals in private firms and when firms that embrace such goals can afford to pursue them by designing workflows that preserve learning opportunities but reduce short-term task efficiency. Contingency factors may include, among others, the level of competition in an industry and the degree to which funders or customers reward organizations that offer learning opportunities to their employees (see Elfenbein & McManus, 2010; European Commission, 2020; Gartenberg et al., 2019). Studies on the role of learning goals in shaping AI adoption could build on related work in strategy and organizations that examines how firms balance profit with other non-financial goals such as sustainability or employment (Ethiraj & Levinthal, 2009; Keum & Liu, 2025; Keum & Meier, 2023; Stevens et al., 2015).

6.3 Contributions

6.3.1 Contributions to the organizational literature

We offer several contributions to the organizational literature on AI. First, much of the existing organizational research treats AI deployment as a shock and studies its downstream effects on important outcomes such as professional identity, learning, or performance (Jia et al., 2023; Kellogg et al., 2020; Lebovitz et al., 2022; Sundaresan & Guler, 2025). We shift the focus from the impacts of AI to organization's decisions on whether and how to deploy AI in the first place, showing that human-AI architectures can be predicted from the underlying nature of the work. Recognizing that AI deployment will typically be an organizational choice has important implications for research on the impacts of AI: Studies that assign certain human-AI configurations experimentally should consider whether these configurations would also be the ones that are most likely to emerge in organizations making deliberate design choices. Studies that examine the effects of AI in real organizations should consider more explicitly the nature of the task and what it implies for the potential of AI to automate versus augment human workers. This may enable a better understanding of why we observe a particular human-AI architecture in a given case, or why performance and employment effects of AI may differ across cases (Dell'Acqua et al., 2023; Jia et al., 2023; Lebovitz et al., 2022).

Second, we expand the scope of organizational research on AI by integrating the physical dimension of work. While prior scholarship has focused heavily on cognitive knowledge work and decision making, we demonstrate that task physicality is a critical structural determinant of how organizations deploy AI. This is an important extension because most tasks in the broader economy require at least some material agency (Felten et al., 2021). Moreover, we move beyond the conventional dichotomy of cognitive versus physical tasks (Autor et al., 2003) to establish hybrid tasks as a conceptually distinct and empirically relevant category.

While a simple dichotomy may be sufficient to study AI-based automation, the category of hybrid tasks is essential for understanding augmentation because hybrid tasks require a broader diversity of inputs than purely cognitive or physical tasks. As such, hybrid tasks are where the complementarities between human sensorimotor skills and AI's predictive power are most pronounced, leading organizations to design joint workflows where machine intelligence supports rather than substitutes for human labor.[23]

Third, our framework helps theorize future developments as technology improves. Although we follow common definitions of AI by excluding material agency (Agrawal et al., 2018; Felten et al., 2021), organizations could of course try to use robots rather than humans to supply the sensorimotor skills required for physical or hybrid tasks. But there appears to be a fundamental asymmetry in the development of these two types of technologies: While AI's cognitive capabilities are advancing at breathtaking speed, machines' ability to navigate and manipulate noisy, unpredictable physical environments remains highly constrained by the physics and economics of robotics (Autor, 2024; Moravec, 1988). Our framework suggests that this gap will function as a "structural anchor" for augmentation: Even as AI becomes increasingly powerful in the cognitive domain, the high costs and limited performance of robots operating in unconstrained environments suggests that the bundling of human sensorimotor skills with AI-based prediction will often remain the optimal solution for hybrid tasks.[24]

---

[23] Although we focus on given tasks, AI may also enable organizations to substitute tasks for each other. For example, AlphaFold allows scientists to reduce the amount of physical experimentation needed to determine protein structures by performing cognitive tasks much more effectively than humans (Jumper et al., 2021). We thank an anonymous reviewer for raising this interesting point.

[24] Despite impressive advances in high-level cognition, AI often fails at the low-level sensorimotor coordination required to perceive noisy, unpredictable physical environments and adjust motor responses simultaneously – a phenomenon known as Moravec's Paradox (Autor, 2024; Brynjolfsson & Mitchell, 2017; Moravec, 1988). Moreover, each physical robotic unit

Finally, our exploratory analysis suggests that organizations' design of human-AI architectures will also depend on their underlying goals. While applications of the task-based model often assume that all organizations maximize profit and task efficiency – thereby focusing our attention on the mapping between task requirements and AI capabilities – objectives related to learning and human capital development may also play a role. This observation provides a starting point for theorizing why organizations sometimes use the same technology in different structural configurations, even for similar tasks (He et al., 2024; Humlum & Vestergaard, 2025; Kellogg et al., 2020). Strategy and organization scholars are, of course, perfectly equipped to study strategic choice and managerial agency (Ethiraj & Levinthal, 2009; Gartenberg et al., 2019; Keum & Meier, 2023), and they could leverage this expertise to complement the prevalent deterministic perspectives focusing on technological capabilities. Indeed, greater attention to organizational goals and human agency would align well with the recent shift from the question of "What *can* AI do?" to "What *should* AI do?" (Acemoglu, 2023; Acemoglu & Restrepo, 2020; Ellis & Bindley, 2025).

6.3.2 Contributions to the economics of AI and research on AI exposure

Our results also inform research on the economics of AI and labor. First, we provide empirical evidence on the organizational micro-foundations of the task-based model. By focusing our analysis at the individual task level, we provide a clean test of the model's "first stage" – the mapping between task characteristics and the division of labor between humans and AI – while ruling out confounding reinstatement effects (Acemoglu & Restrepo, 2019;

---

creates significant variable costs while the marginal costs of applying AI to cognitive tasks move towards zero. As a result, the augmentation of hybrid tasks is likely to remain more efficient than automation in many contexts.

Autor, 2015). More specifically, our findings regarding task routinization contribute to the debate over whether AI represents a continuation of prior "rules-based" automation technologies (Choudhury et al., 2020; Jia et al., 2023) or a new paradigm where advanced pattern recognition enables the automation of non-routine activities (Aghion et al., 2018; Brynjolfsson & Mitchell, 2017). Supporting the second perspective, we identified several cases where AI is used to automate tasks with low routinization. Overall, however, we find a strong positive relationship between routinization and automation. This suggests that while the mechanism of constraint has likely shifted – from the difficulty of codifying explicit rules to the challenge of gathering sufficient stable data for pattern recognition (Choudhary et al., 2025) – the structural mapping between task routinization and automation remains robust also in the context of AI.

Second, we offer a conceptual contribution by clarifying the divergent conceptualizations of augmentation in organization science versus economics. While economists' focus on productivity and labor market impacts has been extremely fruitful, we argue that integrating the input-based perspective of organizational design can advance research by enabling an "insider econometrics" view of the workflow architectures through which AI substitutes or complements human workers (Brynjolfsson et al., 2023; Ichniowski et al., 1997). Our study suggests that important complementarities may arise not only through the emergence of new tasks (Acemoglu & Restrepo, 2019; Agrawal et al., 2023) but also within a given task, potentially adding to our understanding of when and why exactly AI "augments" humans in the economic sense of increasing their productivity. A better understanding of the links between task characteristics, human-AI architectures, and performance at the level of the task could then inform research about downstream implications that occur across tasks or with respect to labor demand and income shares (what we might call the "second stage" of the task-based model). Such research on task- and organization-level mechanisms may also help

economists understand why recent work on the macro impacts of AI shows diverging and sometimes conflicting results (Acemoglu et al., 2022; Brynjolfsson et al., 2025; Humlum & Vestergaard, 2025).

Finally, we contribute to the emerging research on AI exposure (Acemoglu et al., 2022; Eloundou et al., 2024; Felten et al., 2021; Frey & Osborne, 2017) by shifting the focus from technical potential to realized human-AI architectures. Our results suggest that AI exposure does not necessitate a path towards automation and labor substitution; rather, exposure resulting from different combinations of task routinization and physicality is realized in different human-AI architectures that likely have different implications for the affected workers. Interestingly, many existing exposure measures are already grounded in the mapping between task characteristics and AI capabilities but translate it into a singular measure. Scholars could build on our framework to move beyond estimates of the general AI exposure of different tasks to predict whether this exposure will manifest as automation or as human-AI collaboration. This, in turn, may enable us to better predict shifts in the nature of work and in labor markets, allowing organizations, workers, and policymakers to prepare for both the risks and the opportunities resulting from the rapid advancement of AI.

References

Acemoglu, D. 2023. Panel discussion at AoM Annual Meeting.

Acemoglu, D., & Autor, D. 2011. Skills, tasks and technologies: Implications for employment and earnings, ***Handbook of Labor Economics***, Vol. 4: 1043-1171: Elsevier.

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. 2022. Artificial intelligence and jobs: Evidence from online vacancies. ***Journal of Labor Economics***, 40(S1): S293-S340.

Acemoglu, D., Kong, F., & Restrepo, P. 2025. Tasks at work: comparative advantage, technology and labor demand, ***Handbook of Labor Economics***, Vol. 6: 1-114: Elsevier.

Acemoglu, D., & Restrepo, P. 2018. The race between man and machine: Implications of technology for growth, factor shares, and employment. ***American Economic Review***, 108(6): 1488-1542.

Acemoglu, D., & Restrepo, P. 2019. Automation and new tasks: How technology displaces and reinstates labor. ***Journal of Economic Perspectives***, 33(2): 3-30.

Acemoglu, D., & Restrepo, P. 2020. The wrong kind of AI? Artificial intelligence and the future of labour demand. ***Cambridge Journal of Regions, Economy and Society***, 13(1): 25-35.

Aghion, P., Jones, B. F., & Jones, C. I. 2018. Artificial intelligence and economic growth, ***The Economics of Artificial Intelligence: An Agenda***: 237-282: University of Chicago Press.

Agrawal, A., Gans, J., & Goldfarb, A. 2018. ***Prediction Machines: The Simple Economics of Artificial Intelligence***: Harvard Business Press.

Agrawal, A., Gans, J. S., & Goldfarb, A. 2023. Do we want less automation? ***Science***, 381(6654): 155-158.

Allen, R., & Choudhury, P. 2022. Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. ***Organization Science***, 33(1): 149-169.

Andreasson, J. O., Gotrik, M. R., Wu, M. J., Wayment-Steele, H. K., Kladwang, W., et al. 2022. Crowdsourced RNA design discovers diverse, reversible, efficient, self-contained molecular switches. ***Proceedings of the National Academy of Sciences***, 119(18): e2112979119.

Angrist, J. D., & Pischke, J.-S. 2008. ***Mostly Harmless Econometrics: An Empiricist's Companion***: Princeton University Press.

Autor, D. 2024. Applying AI to rebuild middle class jobs, ***NBER Working Paper #32140***.

Autor, D. H. 2015. Why are there still so many jobs? The history and future of workplace automation. ***Journal of Economic Perspectives***, 29(3): 3-30.

Autor, D. H., Levy, F., & Murnane, R. J. 2003. The skill content of recent technological change: An empirical exploration. ***The Quarterly Journal of Economics***, 118(4): 1279-1333.

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

Balázs, B., Mooney, P., Nováková, E., Bastin, L., & Arsanjani, J. J. 2021. Data quality in citizen science. ***The science of citizen science***, 139.

Beane, M. 2019. Shadow learning: Building robotic surgical skill when approved means fail. ***Administrative Science Quarterly***, 64(1): 87-123.

Boussioux, L., N Lane, J., Zhang, M., Jacimovic, V., & Lakhani, K. R. 2024. The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing. ***Organization Science***, 35: 1589-1607.

Brynjolfsson, E., Chandar, B., & Chen, R. 2025. Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence, ***Working Paper***.

Brynjolfsson, E., Li, D., & Raymond, L. R. 2023. Generative AI at work, ***NBER Working***

Paper #31161.
Brynjolfsson, E., & Mitchell, T. 2017. What can machine learning do? Workforce implications. ***Science***, 358(6370): 1530-1534.
Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., et al. 2019. Opportunities and risks for citizen science in the age of artificial intelligence. ***Citizen Science: Theory and Practice***, 4(1).
Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. 2025. Human-AI ensembles: When can they work? ***Journal of Management***, 51(2): 536-569.
Choudhury, P., Starr, E., & Agarwal, R. 2020. Machine learning and human capital complementarities: Experimental evidence on bias mitigation. ***Strategic Management Journal***, 41(8): 1381-1411.
Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., et al. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. ***Working Paper***.
Elfenbein, D. W., & McManus, B. 2010. A greater price for a greater good? Evidence that consumers pay more for charity-linked products. ***American Economic Journal: Economic Policy***, 2(2): 28-60.
Ellis, L., & Bindley, K. 2025. AI is wrecking an already fragile job market for college graduates. ***The Wall Street Journal***.
Eloundou, T., Manning, S., Mishkin, P., & Rock, D. 2024. GPTs are GPTs: Labor market impact potential of LLMs. ***Science***, 384(6702): 1306-1308.
Engelbart, D. C. 1962. Augmenting human intellect: A conceptual framework: Airforce Office of Scientific Research.
Ethiraj, S. K., & Levinthal, D. 2009. Hoping for A to Z while rewarding only A: Complex organizations and multiple goals. ***Organization Science***, 20(1): 4-21.
European Commission. 2020. Citizen science and citizen engagement: European Commission.
Felten, E., Raj, M., & Seamans, R. 2021. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. ***Strategic Management Journal***, 42(12): 2195-2217.
Fortson, L., Crowston, K., Kloetzer, L., & Ponti, M. 2024. Artificial intelligence and the future of citizen science. ***Citizen Science: Theory and Practice***, 9(1).
Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., et al. 2019. Toward understanding the impact of artificial intelligence on labor. ***Proceedings of the National Academy of Sciences***, 116(14): 6531-6539.
Franzoni, C., Poetz, M., & Sauermann, H. 2022. Crowds, citizens, and science: A multi-dimensional framework and agenda for future research. ***Industry and Innovation***, 29(2).
Franzoni, C., & Sauermann, H. 2014. Crowd Science: The organization of scientific research in open collaborative projects. ***Research Policy***, 43(1): 1-20.
Frey, C. B., & Osborne, M. A. 2017. The future of employment: How susceptible are jobs to computerisation? ***Technological Forecasting and Social Change***, 114: 254-280.
Fügener, A., Walzner, D. D., & Gupta, A. 2026. Roles of artificial intelligence in collaboration with humans: Automation, augmentation, and the future of work. ***Management Science***, 72(1): 538-557.
Garicano, L., & Rayo, L. 2025. Training in the age of AI: A theory of apprenticeship viability, ***Working Paper***.
Gartenberg, C., Prat, A., & Serafeim, G. 2019. Corporate purpose and financial performance. ***Organization Science***, 30(1): 1-18.
Hackman, J. R., & Oldham, G. R. 1976. Motivation through the design of work - Test of a

theory. ***Organizational Behavior and Human Performance***, 16(2): 250.
Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., et al. 2021. What Is Citizen Science? The challenges of definition. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perello, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), ***The Science of Citizen Science***: 13-33: Springer.
He, V. F., Shrestha, Y. R., Puranam, P., & Miron-Spektor, E. 2024. Creative together: Human-AI ensembling for idea generation, ***INSEAD Working Paper***.
Hecker, S., Garbe, L., & Bonn, A. 2018. The European citizen science landscape – a snapshot. In S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn (Eds.), ***Citizen Science: Innovation in Open Science, Society and Policy***: 190-200. London: UCL Press.
Huang, M.-H., & Rust, R. T. 2018. Artificial intelligence in service. ***Journal of Service Research***, 21(2): 155-172.
Humlum, A., & Vestergaard, E. 2025. Large language models, small labor market effects, ***NBER Working Paper #33777***.
Ichniowski, C., Shaw, K., & Prennushi, G. 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. ***American Economic Review***, 87(3): 291-313.
Jia, N., Luo, X., Fang, Z., & Liao, C. 2023. When and how artificial intelligence augments employee creativity. ***Academy of Management Journal***.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., et al. 2021. Highly accurate protein structure prediction with AlphaFold. ***Nature***, 596(7873): 583-589.
Kelling, S., Gerbracht, J., Fink, D., Lagoze, C., Wong, W.-K., et al. 2013. A human/computer learning network to improve biodiversity conservation and research. ***AI Magazine***, 34(1): 10-20.
Kellogg, K. C., Valentine, M. A., & Christin, A. 2020. Algorithms at work: The new contested terrain of control. ***Academy of Management Annals***, 14(1): 366-410.
Keum, D. D., & Liu, X. L. 2025. Managerial prosocial preferences and guilt as an emotional barrier to exit decisions. ***Management Science***.
Keum, D. D., & Meier, S. 2023. License to layoff? Unemployment insurance and the moral cost of layoffs. ***Organization Science***.
Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., et al. 2011. Algorithm discovery by protein folding game players. ***Proceedings of the National Academy of Sciences***, 108(47): 18949-18953.
Kittur, A., Yu, L., Hope, T., Chan, J., Lifshitz-Assaf, H., et al. 2019. Scaling up analogical innovation with crowds and AI. ***Proceedings of the National Academy of Sciences***, 116(6): 1870-1877.
Koehler, M., & Sauermann, H. 2024. Algorithmic management in scientific research. ***Research Policy***, 53(4): 104985.
Lauenroth, A., Ioannidis, A. E., & Teichmann, B. 2016. Influence of combined physical and cognitive training on cognition: a systematic review. ***BMC Geriatrics***, 16: 1-14.
Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. 2022. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. ***Organization Science***, 33(1): 126-148.
Liu, Z., Tieu, A., Patel, N., Soultanidis, G., Deyer, L., et al. 2024. ***VIS-MAE: An efficient self-supervised learning approach on medical image segmentation and classification***. Paper presented at the International Workshop on Machine Learning in Medical Imaging.
Luo, S., Li, X., Wang, D., Li, J., & Sun, C. 2015. ***Automatic fish recognition and counting in***

video footage of fishery operations. Paper presented at the 2015 International Conference on Computational Intelligence and Communication Networks (CICN).
Lyons, E., & Zhang, L. 2019. Trade-offs in motivating volunteer effort: Experimental evidence on voluntary contributions to science. ***PLOS ONE***, 14(11).
McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., et al. 2024. AI adoption in America: Who, what, and where. ***Journal of Economics & Management Strategy***, 33(2): 375-415.
Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. 2021. Algorithmic management of work on online labor platforms: When matching meets control. ***MIS Quarterly***, 45(4).
Moravec, H. 1988. ***Mind children: The future of robot and human intelligence***: Harvard University Press.
National Academies. 2018. ***Learning through citizen science: Enhancing opportunities by design***: National Academies Press.
Pachidi, S., Berends, H., Faraj, S., & Huysman, M. 2021. Make way for the algorithms: Symbolic actions and change in a regime of knowing. ***Organization Science***, 32(1): 18-41.
Parrish, J. K., Jones, T., Burgess, H. K., He, Y., Fortson, L., et al. 2019. Hoping for optimality or designing for inclusion: Persistence, learning, and the social network of citizen science. ***Proceedings of the National Academy of Sciences***, 116(6): 1894-1901.
Pentland, B. T., & Rueter, H. H. 1994. Organizational routines as grammars of action. ***Administrative Science Quarterly***: 484-510.
Poetz, M., & Sauermann, H. 2024. ***How and When to Involve Crowds in Scientific Research***. Cheltenham, UK: Edward Elgar.
Ponti, M., & Seredko, A. 2022. Human-machine-learning integration and task allocation in citizen science. ***Humanities and Social Sciences Communications***, 9(1): 1-15.
Puranam, P. 2018. ***The Microstructure of Organizations***: Oxford University Press.
Puranam, P. 2021. Human–AI collaborative decision-making as an organization design problem. ***Journal of Organization Design***, 10(2): 75-80.
Puranam, P., Alexy, O., & Reitzig, M. 2014. What's "new" about new forms of organizing? ***Academy of Management Review***, 39(2): 162-180.
Raisch, S., & Fomina, K. 2023. Combining human and artificial intelligence: Hybrid problem-solving in organizations. ***Academy of Management Review***, 50(2): 441-464.
Raisch, S., & Krakowski, S. 2021. Artificial intelligence and management: The automation–augmentation paradox. ***Academy of Management Review***, 46(1): 192-210.
Sahiner, B., Chen, W., Samala, R. K., & Petrick, N. 2023. Data drift in medical machine learning: implications and potential remedies. ***The British Journal of Radiology***, 96(1150): 20220878.
Sauermann, H., & Franzoni, C. 2015. Crowd science user contribution patterns and their implications. ***Proceedings of the National Academy of Sciences***, 112(3): 679-684.
Shibayama, S., Baba, Y., & Walsh, J. P. 2015. Organizational design of university laboratories: Task allocation and lab performance in Japanese bioscience laboratories. ***Research Policy***, 44(3): 610-622.
Stevens, R., Moray, N., Bruneel, J., & Clarysse, B. 2015. Attention allocation to multiple goals: The case of for-profit social enterprises. ***Strategic Management Journal***, 36(7): 1006-1016.
Sundaresan, S., & Guler, I. 2025. Algorithmic recommendation tools and experiential learning in clinical care. ***Organization Science***, 36(5): 1786-1802.
Svanberg, M., Li, W., Fleming, M., Goehring, B., & Thompson, N. 2024. Beyond AI exposure:

Which tasks are cost-effective to automate with computer vision?, ***SSRN Working Paper #4700751***.

Sytsma, T., & Sousa, É. M. 2023. Artificial Intelligence and the Labor Force: A Data-Driven Approach to Identifying Exposed Occupations: RAND Corporation.

Trouille, L., Lintott, C. J., & Fortson, L. F. 2019. Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. ***Proceedings of the National Academy of Sciences***, 116(6): 1902-1909.

Tschang, F. T., & Almirall, E. 2021. Artificial intelligence as augmenting automation: Implications for employment. ***Academy of Management Perspectives***, 35(4): 642-659.

Wellington-Oguri, R., Fisker, E., Zada, M., Wiley, M., & Townley, J. 2020. Evidence of an unusual poly (A) RNA signature detected by high-throughput chemical mapping. ***Biochemistry***, 59(22): 2041-2046.

Zevin, M., Jackson, C. B., Doctor, Z., Wu, Y., Østerlund, C., et al. 2024. Gravity Spy: lessons learned and a path forward. ***The European Physical Journal Plus***, 139(1): 100.

Table 1: Three human-AI architectures

	Definition	**Examples**
Human-only architecture	All task instances are allocated to humans – no involvement of AI.	• Baseline case. E.g., human agents make all customer service calls without involvement of AI, humans make image classifications without involvement of AI.
Automation architecture	Task instances are allocated to AI for independent execution, without human involvement in those instances (independent AI). • Full automation: all task instances performed by AI. • Partial automation: only a subset of task instances is performed by AI.	• AI performs all task instances without human involvement. E.g., a pre-trained algorithm classifies all medical images for a hospital (Liu et al., 2024). • Humans in the organization generate training data by working on some task instances, AI then performs remaining task instances. E.g., classifications of galaxies in *Galaxy Zoo* (Trouille et al., 2019).
Augmentation architecture	The same task instances are allocated to humans and AI for joint execution, bundling human and AI inputs (joint human+AI). • Full augmentation: all task instances performed jointly by humans and AI. • Partial augmentation: only a subset of task instances is performed jointly by humans and AI.	• AI provides recommendations about the same task instance to the human. E.g., classification of medical images in Lebovitz et al. (2022), identification of birds in *eBird*, and support during customer service calls in Brynjolfsson et al. (2023). • AI and humans make independent predictions for the same task instance, but their predictions are then aggregated to arrive at the final prediction for that particular instance (Puranam, 2021).

Note: The three architectures are collectively exhaustive in that they capture all possible allocations of task instances to humans and/or AI. They are not mutually exclusive in that organizations can use both partial automation and partial augmentation for a given task.

Table 2: Summary statistics

Variable	N	Mean	Min	Max
Automation	1,205	0.08	0	1
Augmentation	1,205	0.09	0	1
Routinization: Low	1,205	0.10	0	1
Routinization: Medium	1,205	0.75	0	1
Routinization: High	1,205	0.15	0	1
Task physicality: Physical	1,205	0.13	0	1
Task physicality: Hybrid	1,205	0.68	0	1
Task physicality: Cognitive	1,205	0.19	0	1
Learning goals	1,205	0.15	0	1
Platform affiliated	1,205	0.41	0	1
Age	1,205	5.13	1	6
Secondary tasks	1,205	0.08	0	1
Geo scope: Local / regional	1,205	0.57	0	1
Geo scope: National / international	1,205	0.31	0	1
Geo scope: Missing	1,205	0.12	0	1
Mainfield: Biology	1,205	0.46	0	1
Mainfield: Earth / environmental science	1,205	0.37	0	1
Mainfield: Physical sciences	1,205	0.06	0	1
Mainfield: Social sciences / humanities	1,205	0.06	0	1
Mainfield: Other	1,205	0.06	0	1
Secondfield: Biology	1,205	0.06	0	1
Secondfield: Earth / environmental sciences	1,205	0.14	0	1
Secondfield: Physical sciences	1,205	0.01	0	1
Secondfield: Social sciences / humanities	1,205	0.02	0	1
Secondfield: Other	1,205	0.03	0	1
Secondfield: None	1,205	0.73	0	1

Table 3: Main regressions of Automation and Augmentation

Variable	1 LPM Autom.	2 LPM Augm.	3 LPM Autom.	4 LPM Augm.	5 LPM Autom.	6 LPM Augm.	7 LPM AnyAI
Routinization: Low	-0.337**	-0.045*			-0.163**	-0.035	-0.097
	[0.043]	[0.023]			[0.054]	[0.022]	[0.059]
Routinization: Medium	-0.350**	Omitted			-0.154**	Omitted	-0.050
	[0.039]				[0.051]		[0.055]
Routinization: High	Omitted	-0.130**			Omitted	-0.120**	Omitted
		[0.023]				[0.035]	
Task physicality: Physical			-0.348**	-0.087**	-0.242**	-0.082**	-0.279**
			[0.035]	[0.012]	[0.045]	[0.012]	[0.049]
Task physicality: Hybrid			-0.366**	Omitted	-0.257**	Omitted	-0.212**
			[0.035]		[0.045]		[0.049]
Task physicality: Cognitive			Omitted	-0.102**	Omitted	-0.019	Omitted
				[0.020]		[0.032]	
Platform affiliated	0.016	0.157**	0.032*	0.143**	0.020+	0.151**	0.171**
	[0.013]	[0.020]	[0.013]	[0.019]	[0.012]	[0.020]	[0.023]
Age	-0.009	-0.006	-0.007	-0.007	-0.006	-0.008	-0.015
	[0.011]	[0.009]	[0.011]	[0.009]	[0.011]	[0.009]	[0.014]
Secondary tasks	0.012	-0.025	-0.016	-0.007	-0.000	-0.010	-0.005
	[0.022]	[0.023]	[0.022]	[0.022]	[0.021]	[0.023]	[0.032]
Geo scope: Local / regional	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Geo scope: National / intern.	0.004	0.009	-0.006	0.002	-0.008	0.003	-0.004
	[0.028]	[0.028]	[0.027]	[0.028]	[0.027]	[0.028]	[0.037]
Geo scope: Missing	-0.014	-0.044	-0.022	-0.046	-0.014	-0.051+	-0.064+
	[0.021]	[0.029]	[0.021]	[0.029]	[0.021]	[0.029]	[0.035]
Mainfield: Biology	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Mainfield: Earth / env. sci.	-0.041**	-0.047*	-0.030**	-0.041*	-0.032**	-0.042*	-0.072**
	[0.012]	[0.020]	[0.011]	[0.020]	[0.011]	[0.020]	[0.023]
Mainfield: Physical sci.	0.191**	-0.020	0.150**	-0.026	0.148**	-0.021	0.108*
	[0.052]	[0.035]	[0.052]	[0.036]	[0.051]	[0.035]	[0.054]
Mainfield: Social sci. / hum.	-0.045	-0.074**	-0.113**	-0.063**	-0.097*	-0.069**	-0.157**
	[0.039]	[0.023]	[0.042]	[0.023]	[0.042]	[0.022]	[0.047]
Mainfield: Other	0.088+	-0.014	0.042	-0.012	0.045	-0.013	0.026
	[0.048]	[0.034]	[0.044]	[0.033]	[0.044]	[0.032]	[0.048]
Secondfield: Biology	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Secondfield: Earth / env. sci.	-0.122**	-0.027	-0.097**	-0.023	-0.098**	-0.024	-0.123*
	[0.034]	[0.039]	[0.030]	[0.040]	[0.030]	[0.040]	[0.049]
Secondfield: Physical sci.	-0.141+	0.015	-0.196*	0.024	-0.171*	0.015	-0.144
	[0.085]	[0.079]	[0.078]	[0.078]	[0.080]	[0.081]	[0.109]
Secondfield: Social sci. / hum.	-0.128*	0.032	-0.164*	0.035	-0.152*	0.030	-0.117
	[0.060]	[0.060]	[0.066]	[0.060]	[0.063]	[0.060]	[0.085]
Secondfield: Other	0.035	-0.023	-0.018	-0.010	0.005	-0.022	-0.036
	[0.065]	[0.046]	[0.057]	[0.049]	[0.057]	[0.046]	[0.066]
Secondfield: None	-0.100**	-0.007	-0.093**	0.002	-0.094**	0.001	-0.095*
	[0.032]	[0.030]	[0.029]	[0.030]	[0.029]	[0.030]	[0.041]
Constant	0.506**	0.119+	0.496**	0.124+	0.534**	0.135*	0.527**
	[0.084]	[0.067]	[0.078]	[0.068]	[0.080]	[0.068]	[0.099]
Observations	1,205	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.372	0.087	0.400	0.087	0.416	0.095	0.228
df	16	16	16	16	18	18	18

Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%. *AnyAI* takes the value of 1 if a project uses automation and/or augmentation. LPM = Linear Probability Model (OLS).

Table 4: Exploring the role of learning goals

Variable	1 LPM Autom.	2 LPM Augm.	3 LPM Autom.	4 LPM Augm.	5 Ologit Routinization	6 Ologit CogHybPhys
Learning goals	-0.044**	0.029	0.001	0.026	-0.896**	-0.922**
	[0.016]	[0.024]	[0.014]	[0.024]	[0.222]	[0.196]
Routinization: Low			-0.163**	-0.037		
			[0.054]	[0.023]		
Routinization: Medium			-0.154**	Omitted		
			[0.051]			
Routinization: High			Omitted	-0.119**		
				[0.035]		
Task physicality: Physical			-0.242**	-0.084**		
			[0.045]	[0.012]		
Task physicality: Hybrid			-0.258**	Omitted		
			[0.045]			
Task physicality: Cognitive			Omitted	-0.016		
				[0.032]		
Controls	Included	Included	Included	Included	Included	Included
Constant	0.278**	0.067	0.534**	0.132*		
	[0.083]	[0.068]	[0.081]	[0.067]		
Observations	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.214	0.069	0.416	0.096		
df	15	15	19	19	15	15

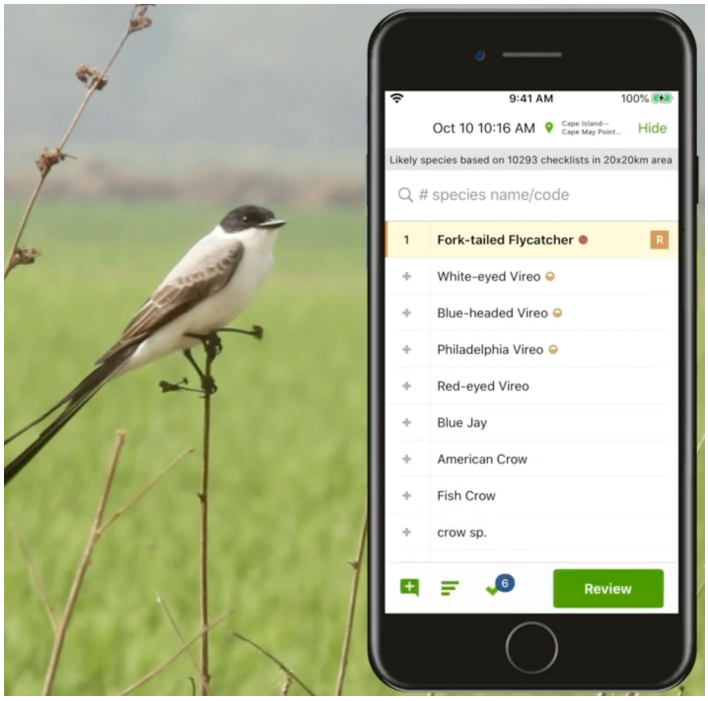
Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%. *CogHybPhys* is the 3-point measure of task physicality (1=physical, 2=hybrid, 3=cognitive). LPM = Linear Probability Model (OLS); Ologit = Ordered logistic regression. Full set of controls as per main regressions reported in Table 3.

Table 5: Robustness checks controlling for project size

Variable	1 LPM Autom.	2 LPM Augm.	3 LPM Autom.	4 LPM Augm.	5 LPM Autom.	6 LPM Augm.
Routinization: Low	-0.143**	-0.031			-0.143**	-0.033
	[0.054]	[0.021]			[0.054]	[0.022]
Routinization: Medium	-0.137**	Omitted			-0.137**	Omitted
	[0.051]				[0.051]	
Routinization: High	Omitted	-0.154**			Omitted	-0.154**
		[0.035]				[0.035]
Task physicality: Physical	-0.227**	-0.069**			-0.226**	-0.071**
	[0.044]	[0.012]			[0.044]	[0.012]
Task physicality: Hybrid	-0.248**	Omitted			-0.248**	Omitted
	[0.044]				[0.044]	
Task physicality: Cognitive	Omitted	-0.026			Omitted	-0.024
		[0.030]				[0.030]
Learning goals			-0.041**	0.025	-0.002	0.017
			[0.016]	[0.023]	[0.014]	[0.023]
Size: <=1,000	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Size: 1,001-10,000	0.055**	0.067*	0.133**	0.034	0.055**	0.067*
	[0.021]	[0.030]	[0.023]	[0.029]	[0.021]	[0.030]
Size: >10,000	0.072**	0.076*	0.178**	0.035	0.072**	0.075*
	[0.027]	[0.036]	[0.033]	[0.035]	[0.027]	[0.036]
Size: Missing	0.011	-0.056**	0.033**	-0.064**	0.011	-0.055**
	[0.010]	[0.018]	[0.012]	[0.018]	[0.010]	[0.018]
Other controls	Included	Included	Included	Included	Included	Included
Constant	0.511**	0.166*	0.266**	0.091	0.511**	0.164*
	[0.080]	[0.068]	[0.081]	[0.070]	[0.080]	[0.068]
Observations	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.424	0.125	0.265	0.089	0.424	0.125
df	21	21	18	18	22	22

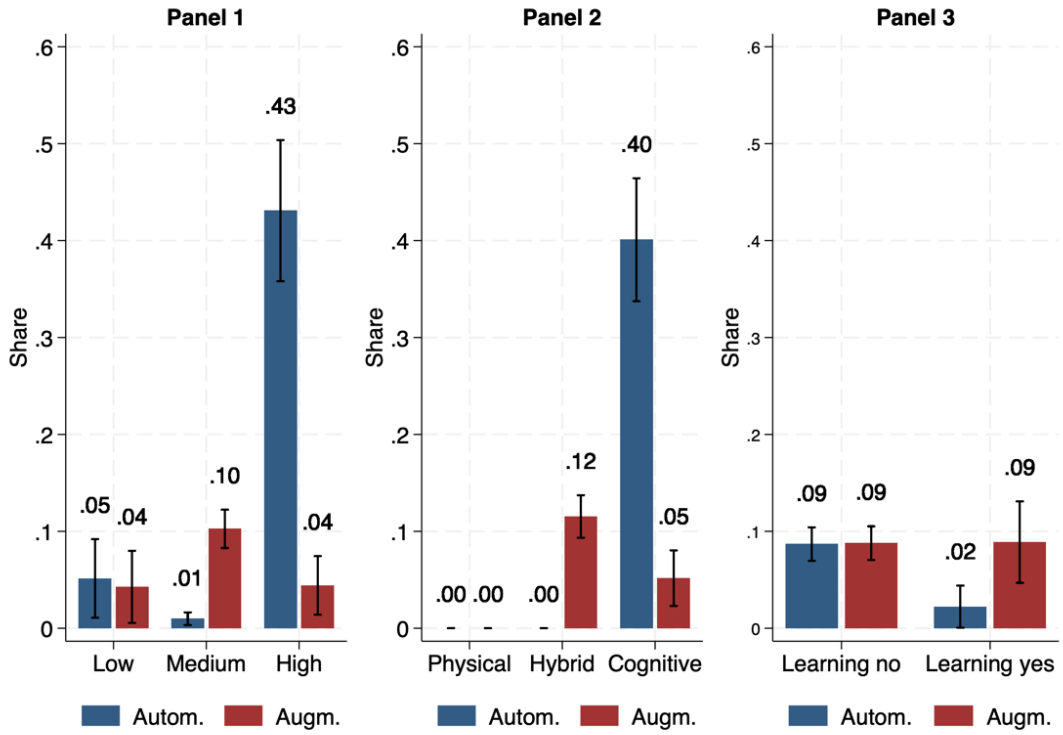
Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%. LPM = Linear Probability Model (OLS). Full set of controls as per main regressions reported in Table 3.

Figure 1: Example of augmentation: AI supports the recording of observations in *eBird*



Source: Cornell Lab of Ornithology, https://www.youtube.com/watch?v=gkN7ml7lMxw

Figure 2: Share of projects using automation and augmentation, by task routinization (Panel 1), physical vs. cognitive nature of the task (Panel 2), and learning goals (Panel 3)



Note: Numbers above the 95% confidence intervals indicate the height of the bar.

The Organizational Design of Human-AI Architectures: A Task-Based Framework and Evidence from Crowd Science Projects

Appendix A: Projects mentioned in the paper

Appendix B: Perspectives on automation and augmentation

Appendix C: Key correlations

Appendix A: Projects mentioned in main paper

Table A1 lists all crowd science projects mentioned in the main paper, with links to their web profiles.

Table A1: Projects with links to web profiles (as of time of data collection)

Project name	**Link**
Aurorasaurus	https://www.aurorasaurus.org/
Bellingen Riverwatch	https://scistarter.org/bellingen-riverwatch-our-rivers-our-future
DeVOTE	https://www.votemeanings.eu/
eBird	https://ebird.org/home
Eterna	https://eternagame.org/
Exoplanet Watch	https://exoplanets.nasa.gov/exoplanet-watch/about-exoplanet-watch/overview/
Foldit	https://fold.it/
Frog ID	https://www.frogid.net.au/
Galaxy Zoo	https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/
Global Fishing Watch	https://globalfishingwatch.org/
Glyph	https://glyph.shh.mpg.de/
Gravity Spy	https://www.zooniverse.org/projects/zooniverse/gravity-spy
Illinois Butterfly Monitoring Network	https://scistarter.org/illinois-butterfly-monitoring-network
Industrial Smoke Hunting	https://smoke.createlab.org/
Justice BRD	https://www.justicebrd.com/
LA Spider Survey	https://scistarter.org/la-spider-survey
LHC@home	https://lhcathome.cern.ch/lhcathome/index.php
Living with Machines	https://livingwithmachines.ac.uk/
Mammoth Site Ice Age Explorers	https://scistarter.org/the-mammoth-site-ice-age-explorers
Monarch Larva Monitoring Project	https://mlmp.org/
Mystic River Herring	https://www.mysticherring.org/
NOVA Labs	https://www.pbs.org/wgbh/nova/labs/
Personal Genome Project	https://www.personalgenomes.org/

Phylo	https://phylo.cs.mcgill.ca/
Quake Catcher Network	https://scistarter.org/quake-catcher-network
Quantum Minds	https://www.scienceathome.org/games/quantum-minds/
Rescue A Reef	https://rescueareef.earth.miami.edu/
Save the Tasmanian Devil	https://scistarter.org/save-the-tasmanian-devil
Snapshot Safari	https://www.zooniverse.org/organizations/meredithspalmer/snapshot-safari
Stall Catchers	https://stallcatchers.com/main

Appendix B: Perspectives on Automation and Augmentation

Although the terms *automation* and *augmentation* are widely used in discussions about the role and impact of new technologies such as AI, they are not consistently defined and sometimes used with different connotations (Agrawal et al., 2023; Autor, 2015; Jarrahi, 2018; Raisch & Krakowski, 2021). Moreover, it is not always clear how automation and augmentation relate to similar constructs such as displacement effects, reinstatement effects, and productivity effects, as well as substitution and complementarity. Based on our reading of the relevant literature, this appendix briefly discusses the use of automation and augmentation from different perspectives, highlights connections to related constructs, and positions the definitions used in our paper within the broader literature.

Scholars studying the interplay between new technologies (such as AI) and human workers emphasize different aspects along two key dimensions:

- While some prior studies focus on the different *roles* that technologies and human workers play in performing work (input-based), others focus on the *impacts* of those roles on productivity or employment.
- Studies look at these issues at *different levels of analysis* – from individual tasks to entire jobs or occupations to labor in the aggregate.

Of course, a given study may employ multiple perspectives, e.g., by studying both the division of labor between AI and humans as well as its impacts, or by studying the effects of AI adoption at different levels of analysis. Nevertheless, it is useful to distinguish these dimensions to better understand the diversity of approaches and the different assumptions or connotations that scholars may have when using the terms automation and augmentation. In the following, we lay out a simple framework that highlights these differences. Table A2 provides a summary.

Focus on roles versus impacts. Many scholars, especially in the domains of management and human-computer-interaction, seek to understand the *roles* that AI and humans play in performing tasks or workflows, and how their inputs combine (or not) to complete the work. For example, studies have examined how doctors use AI in medical diagnosis, how humans and AI interact in decision making, how digital platforms sequence workflows between humans and AI, or how researchers use AI to perform different scientific tasks (Jarrahi, 2018; Lebovitz et al., 2022; Parasuraman et al., 2000; Ponti & Seredko, 2022; Wang et al., 2023). This work also discusses issues of control, e.g., whether final decisions are made by AI or humans (e.g., “humans in the loop”). In this perspective, the terms automation and

augmentation are defined by the nature of the interaction and the division of labor between humans and the technology rather than by performance outcomes. Automation generally means that AI takes over certain tasks that humans would otherwise do, whereas augmentation means that humans use AI as a tool or that humans and AI jointly work on a task (Lebovitz et al., 2022; Raisch & Krakowski, 2021).

Other scholars, especially in economics, acknowledge that humans and the new technology can interact in different roles but they focus on the *impacts* that these interactions have on productivity and employment (Acemoglu & Restrepo, 2019; Agrawal et al., 2023; Felten et al., 2021; Frank et al., 2019; Tschang & Almirall, 2021). In this perspective, the terms automation and augmentation can carry an outcome connotation: Automation is sometimes assumed to reduce the need for human labor, while augmentation is assumed to increase human productivity and performance (Brynjolfsson, 2022). Scholars taking this perspective distinguish several specific effects that a new technology can have, including *displacement effects* (the demand for labor decreases due to AI substituting for labor), *productivity effects* (if automation reduces costs per unit, then the demand for the product – and thus demand for labor – can increase), as well as *reinstatement effects* (the new technology and declining labor demand in old tasks give rise to new tasks in the same or different jobs) (Acemoglu & Restrepo, 2019). Related concepts are those of *substitution* (i.e., AI and human labor are substitutes in the production function such that AI can automate human tasks) as well as *complementarity* (i.e., technology increases the marginal productivity of human labor).

The distinction between roles and impacts matters because the same roles can have different impacts under different conditions and for different people. For example, AI assisting humans in a task might increase output quality in some cases, but could reduce quality in others if humans rely on AI too much, use AI inappropriately, lack complementary skills, or reduce their own effort (Agarwal et al., 2023; Allen & Choudhury, 2022; Dell'Acqua et al., 2023; Lebovitz et al., 2022). Similarly, automating a task will typically reduce labor demand if output demand is fixed; however, if lower costs from automation stimulate higher demand, then total human labor employed can also increase (i.e., a productivity effect may outweigh a displacement effect) (Acemoglu & Restrepo, 2019). Relatedly, the collaboration of humans and AI in a particular task would be called augmentation by scholars focusing on roles, but this collaboration could reduce (rather than increase) demand for labor if output demand is limited and can be satisfied by fewer but more efficient workers.

In the main paper, *we follow the "roles" perspective* by defining augmentation and automation with respect to the roles that AI and humans play in performing tasks. We do not assume any inherent positive or negative effects of automation and augmentation on outcomes such as productivity and labor demand. Studying the effects of automation and augmentation on such outcomes requires different types of data and remains an important area for future research.

Different levels of analysis. Researchers also differ in the level of analysis at which they study the roles and impacts of new technologies such as AI. We highlight three common levels:

Task level. Tasks are specific work activities or units of work, such as classifying medical images, analyzing a business case, or making phone calls to customers. A large stream of research focuses on which tasks can be automated or augmented by a new technology, often considering underlying task characteristics such as routinization, the physical versus cognitive nature of tasks, or the need for human interactions (Autor et al., 2003; Brynjolfsson & Mitchell, 2017). If using the roles perspective at this level, scholars often examine how humans and AI interact in a workflow (Dell'Acqua et al., 2023; Ponti & Seredko, 2022). If using the impacts perspective, they may examine how performing a task with AI assistance affects the quality or quantity of output, or the time spent by human workers (Allen & Choudhury, 2022; Brynjolfsson et al., 2023; Choudhury et al., 2020; Jia et al., 2023).

Occupation/job level. Other research focuses on entire jobs or occupations, which are assumed to entail multiple tasks. Studies using the roles perspectives at this level may ask when organizations use AI for jobs previously done by human workers or how workers in an occupation incorporate AI tools in their workflows (Tschang & Almirall, 2021; Wang et al., 2023). Scholars using the impacts perspective may study questions such as how the introduction of new technologies has changed the share of different occupations in an organization or in the labor force, or the wages earned in those occupations (Autor & Handel, 2013; Brynjolfsson et al., 2018). Some of this work is grounded in arguments about the task content of different occupations, but empirical analyses are then performed at the level of occupations by aggregating across tasks (Eloundou et al., 2024; Felten et al., 2021; Frey & Osborne, 2017). This level of analysis also enables scholars to study how the composition of tasks has changed within a given occupation; for example, new tasks may emerge as old tasks are automated, preserving employment at the level of the occupation due to a reinstatement effect (Frank et al., 2019).

Aggregate labor level. A third level of analysis considers labor in a broad, aggregated sense, often contrasted with capital. Among others, work at this level explores how the labor share of income has changed over time and in response to new technologies, perhaps also studying how effects differ in different parts of the ability distribution (Acemoglu & Restrepo, 2018; Dixon et al., 2021). These studies may be based on theories about tasks or occupations (e.g., that highly educated people have jobs that entail more complex tasks, which are more difficult to automate), but the outcomes are observed at more aggregate levels. Such work may also study changes in occupational structures, e.g., that a new technology reduces employment in some occupations but gives rise to new occupations (i.e., a reinstatement effect) (Autor, 2015; Autor et al., 2003). At this high level of aggregation, investigations of the specific roles of humans and AI are difficult, and researchers tend to focus on impacts.

The distinction between different levels of analysis matters because the same technology can interact differently with humans when analyzed at different levels of analysis. For example, AI may automate a particular task A (task-level perspective) but may, by doing so, increase the productivity of a worker performing a range of complementary tasks B and C (occupation-level perspective aggregating across tasks) (Raisch & Krakowski, 2021). Relatedly, AI may have different impacts in terms of labor demand, e.g., reducing the need for labor in a task A that is automated while increasing the demand for workers in an occupation that performs complementary tasks D and E (Bessen, 2019). Distinguishing different levels of analysis can also help resolve apparent paradoxes. For example, the common notion that "one person's automation is another's augmentation" (Agrawal et al., 2023) is less confusing when we realize that it typically refers to persons working in two different occupations (e.g., AI tools that automate the work of a management assistant may augment a manager).

In the main paper, *we focus on the task level of analysis*. As such, we theorize about the relationships between task characteristics and the automation vs. augmentation of individual tasks, and we draw on unique task-level data to test our predictions. We do not consider effects at higher levels of aggregation (e.g., occupations or aggregate labor). We also do not study effects that play out across tasks or occupations (e.g., reinstatement effects). This focus enables us to build on the task-based model (Acemoglu & Autor, 2011) to analyze organizational decisions to automate versus augment a particular task and avoids complexities that arise from the interplay between different effects at higher levels of analysis.

Table A2: Different perspectives on the roles and impacts of technology using different levels of analysis; resulting differences in the conceptualizations or connotations of *automation* and *augmentation*.

	Focus on *Roles*	**Focus on *Impacts***
Task level	Typical questions: When do firms automate tasks versus keep humans in the loop? How do AI and human collaborate in a task? *Automation*: Conceptualized as AI taking over a task or task instances. *Augmentation*: Conceptualized as AI and humans collaborating on a task or on task instances.	Typical questions: How does using AI affect the time spent by humans on the task? How does using AI affect the quality or quantity of task outputs? *Automation*: AI often assumed to reduce the labor required for the task. *Augmentation*: AI often assumed to increase the productivity of humans in the task.
Occupation/ job level	Typical questions: Which occupations are adopting AI, and how? How does the task content of occupations change due to AI? *Automation*: Conceptualized as AI performing the work previously performed by humans in an occupation. *Augmentation*: Conceptualized as workers in an occupation using AI as a tool or collaborator.	Typical questions: Which occupations are most "exposed" to the risk of job loss due to automation? How does using AI affect wages for an occupation? Is labor demand in an occupation preserved due to the emergence of new tasks? *Automation*: AI often assumed to reduce demand for labor in an occupation. *Augmentation*: AI often assumed to increase productivity of workers in an occupation.
Aggregate labor level	Level too aggregate to meaningfully study different roles of humans and AI.	Typical questions: What is the effect of AI on total employment? What is the effect of AI on income shares going to labor versus capital? Does AI give rise to new occupations? *Automation*: AI often assumed to reduce demand for labor. *Augmentation*: AI often assumed to increase the productivity of labor, potentially increasing labor demand.

Appendix C: Key correlations

Table A3 reports correlations between key variables.

Table A3: Key correlations

	Variable	1	2	3	4	5	6	7	8	9
1	Automation	1								
2	Augmentation	-0.0459	1							
3	Routinization: Low	-0.0318	-0.0524	1						
4	Routinization: Medium	-0.4397*	0.0897*	-0.5721*	1					
5	Routinization: High	0.5573*	-0.0650*	-0.1379*	-0.7335*	1				
6	Task physicality: Physical	-0.1123*	-0.1206*	0.1134*	0.0460	-0.1496*	1			
7	Task physicality: Hybrid	-0.4181*	0.1397*	-0.0966*	0.5244*	-0.5532*	-0.5616*	1		
8	Task physicality: Cognitive	0.5922*	-0.0625*	0.0176	-0.6616*	0.7844*	-0.1897*	-0.7059*	1	
9	Learning goals	-0.0863*	0.0014	0.1535*	-0.0026	-0.1241*	0.1269*	-0.0037	-0.1042*	1
10	Platform affiliated	0.1663*	0.2386*	-0.0769*	-0.1609*	0.2580*	-0.1458*	-0.0460	0.1794*	-0.0856*
11	Age	-0.2200*	-0.0272	-0.0164	0.2636*	-0.3047*	0.0703*	0.2247*	-0.3268*	0.0296
12	Secondary tasks	-0.0299	-0.0502	0.2285*	-0.0877*	-0.0835*	0.1792*	-0.1225*	-0.0081	0.1968*
13	Geo scope: Local / regional	-0.1828*	-0.0268	-0.0045	0.2182*	-0.2597*	0.1182*	0.1565*	-0.2869*	0.0434
14	Geo scope: National / international	0.2570*	0.0329	-0.0013	-0.2944*	0.3566*	-0.1165*	-0.2466*	0.3924*	-0.0036
15	Geo scope: Missing	-0.0874*	-0.0060	0.0088	0.0866*	-0.1119*	-0.0143	0.1127*	-0.1215*	-0.0611*
16	Mainfield: Biology	-0.1046*	0.0721*	-0.0326	0.1351*	-0.1362*	-0.0426	0.1897*	-0.1886*	-0.0876*
17	Mainfield: Earth / environmental sciences	-0.1418*	-0.0347	-0.0915*	0.1670*	-0.1259*	0.1139*	0.0752*	-0.1867*	0.0593*
18	Mainfield: Physical sciences	0.3238*	-0.0031	0.0489	-0.2650*	0.2795*	-0.0763*	-0.2111*	0.3158*	-0.0356
19	Mainfield: Social sciences / humanities	0.0321	-0.0536	0.1538*	-0.2126*	0.1292*	-0.0046	-0.2073*	0.2499*	0.1006*
20	Mainfield: Other	0.1584*	-0.0252	0.0534	-0.1516*	0.1388*	-0.0630*	-0.1383*	0.2181*	-0.0016

Note: Pairwise Pearson correlations. *=Significant at 5%.

Appendix References

Acemoglu, D., & Autor, D. 2011. Skills, tasks and technologies: Implications for employment and earnings, ***Handbook of Labor Economics***, Vol. 4: 1043-1171: Elsevier.

Acemoglu, D., & Restrepo, P. 2018. The race between man and machine: Implications of technology for growth, factor shares, and employment. ***American Economic Review***, 108(6): 1488-1542.

Acemoglu, D., & Restrepo, P. 2019. Automation and new tasks: How technology displaces and reinstates labor. ***Journal of Economic Perspectives***, 33(2): 3-30.

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. 2023. Combining human expertise with artificial intelligence: Experimental evidence from radiology: National Bureau of Economic Research.

Agrawal, A., Gans, J. S., & Goldfarb, A. 2023. Do we want less automation? ***Science***, 381(6654): 155-158.

Allen, R., & Choudhury, P. 2022. Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. ***Organization Science***, 33(1): 149-169.

Autor, D. H. 2015. Why are there still so many jobs? The history and future of workplace automation. ***Journal of Economic Perspectives***, 29(3): 3-30.

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

Autor, D. H., Levy, F., & Murnane, R. J. 2003. The skill content of recent technological change: An empirical exploration. ***The Quarterly Journal of Economics***, 118(4): 1279-1333.

Bessen, J. 2019. Automation and jobs: When technology boosts employment. ***Economic Policy***, 34(100): 589-626.

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

Brynjolfsson, E., Li, D., & Raymond, L. R. 2023. Generative AI at work, ***NBER Working Paper #31161***.

Brynjolfsson, E., & Mitchell, T. 2017. What can machine learning do? Workforce implications. ***Science***, 358(6370): 1530-1534.

Brynjolfsson, E., Mitchell, T., & Rock, D. 2018. ***What can machines learn and what does it mean for occupations and the economy?*** Paper presented at the AEA papers and proceedings.

Choudhury, P., Starr, E., & Agarwal, R. 2020. Machine learning and human capital complementarities: Experimental evidence on bias mitigation. ***Strategic Management Journal***, 41(8): 1381-1411.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., et al. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. ***Working Paper***.

Dixon, J., Hong, B., & Wu, L. 2021. The robot revolution: Managerial and employment consequences for firms. ***Management Science***, 67(9): 5586-5605.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. 2024. GPTs are GPTs: Labor market impact potential of LLMs. ***Science***, 384(6702): 1306-1308.

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

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., et al. 2019. Toward understanding the impact of artificial intelligence on labor. ***Proceedings of the National Academy of Sciences***, 116(14): 6531-6539.

Frey, C. B., & Osborne, M. A. 2017. The future of employment: How susceptible are jobs to computerisation? ***Technological Forecasting and Social Change***, 114: 254-280.

Jarrahi, M. H. 2018. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. ***Business Horizons***, 61(4): 577-586.

Jia, N., Luo, X., Fang, Z., & Liao, C. 2023. When and how artificial intelligence augments employee creativity. ***Academy of Management Journal***.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. 2022. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. ***Organization Science***, 33(1): 126-148.

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

Ponti, M., & Seredko, A. 2022. Human-machine-learning integration and task allocation in citizen science. ***Humanities and Social Sciences Communications***, 9(1): 1-15.

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

Tschang, F. T., & Almirall, E. 2021. Artificial intelligence as augmenting automation: Implications for employment. ***Academy of Management Perspectives***, 35(4): 642-659.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., et al. 2023. Scientific discovery in the age of artificial intelligence. ***Nature***, 620(7972): 47-60.