Automation or Augmentation?
Task Characteristics, Human-Centered Goals, and the Deployment of AI

Maximilian Koehler
ESMT Berlin, maximilian.koehler@phd.esmt.org

Henry Sauermann
ESMT Berlin, henry.sauermann@esmt.org

Abstract

The rise of artificial intelligence raises the question of when AI is adopted and whether it is deployed to replace versus augment workers. Studies of prior technologies such as robots and computers emphasize the role of task routinization and the cognitive vs. physical nature of tasks, yet theorizing and empirical evidence in the context of AI technologies are scarce. Moreover, the diffusion of AI is not automatic: Decisions on whether and how to use AI are made by organizations, and these organizations may pursue not only production efficiency but also goals that require the continued involvement of humans (e.g., providing jobs, human capital development, community engagement). We develop hypotheses regarding the role of task characteristics and human-centered goals and test them using data on more than 1,200 projects that involve participants in scientific research. Although task characteristics predict AI exposure in a general sense, they relate differently to automation and augmentation, emphasizing the need to distinguish different types of exposure in future work. Projects pursuing human-centered goals are less likely than others to use automation, while they are just as likely to use AI to augment. We contribute to the literature on AI in organizations by providing novel evidence from task-level data and by highlighting that the impact of AI on humans is, partly, a matter of human choice.

Key words: Augmentation, automation, task routinization, human capital, crowd science

Acknowledgements. We thank SciStarter.org for providing us with access to data and ESMT Berlin for financial support. We are grateful for helpful feedback from colleagues, seminar participants, and conference discussants. All errors are our own.

1 Introduction

Artificial intelligence (AI) applications are diffusing rapidly across organizations in many sectors (Agrawal et al., 2018; Bianchini et al., 2022; Mollick & Mollick, 2023; Raisch & Krakowski, 2021). Echoing concerns about the impact of prior technologies such as robots and computers, this development has triggered discussions about whether AI will diminish the role of humans by automating tasks, or whether AI will augment humans and make them more effective in performing their tasks. Although some authors claim one side of the debate or the other, more nuanced discussions recognize that the effects will differ across jobs and occupations. The underlying premise is that the strengths and weaknesses of AI – and those of humans – depend on the nature of the task, especially the degree of routinization and the physical vs. cognitive nature of work (Autor, 2015; Brynjolfsson & Mitchell, 2017; Felten et al., 2021). Given that the capabilities and constraints of AI differ from those of prior automation technologies, however, specific predictions differ and there is little empirical evidence on the role of task characteristics in shaping whether and how AI is used.

Perhaps more importantly, decisions regarding the deployment of AI are still made by human decision-makers in organizations. These decision-makers may consider not only potential cost savings or quality improvements from employing AI for a given task; they may also pursue goals that require the continued active involvement of humans (henceforth, "human-centered goals"). Among others, such goals may relate to political pressures to preserve employment (Gu et al., 2020), prosocial preferences to provide jobs (Guenzel et al., 2023; Keum & Meier, 2023), giving workers opportunities to develop human capital (Battilana & Lee, 2014; Frank et al., 2019), and strengthening ties with local communities (Cooper et al., 2021; Marquis & Battilana, 2009). Such human-centered goals have been shown to matter for other managerial decisions but they have received little attention in the context of AI adoption.

We make two key contributions. First, we provide unique empirical evidence on the relationships between task characteristics and the deployment of AI. Prior efforts in this direction relied on an indirect approach: Scholars have assessed the potential "exposure" that occupations have to AI by estimating the overlap between evolving AI capabilities and task requirements or by asking raters to assess how much time AI could save in performing particular tasks (Eloundou et al., 2024; Felten et al., 2021; Sytsma & Sousa, 2023). Although this approach reveals important differences in AI exposure across tasks, it does not tell us whether organizations actually employ AI for tasks that are predicted to have high exposure. More importantly, it does not tell us whether they do so by deploying AI to automate versus augment, which has important implications for the affected people. We contribute more direct evidence by studying a range of tasks performed in real organizations, allowing us to measure task characteristics as well as whether AI is used to automate, to augment, or not used at all. Our data come from more than 1,200 crowd science projects that involve participants in research activities such as data collection, data analysis, or creative problem solving (Franzoni & Sauermann, 2014; Haklay et al., 2021).[1] Each project involves humans in one primary task, but projects collectively span a broad range of tasks that differ with respect to their degree of routinization as well as their cognitive vs. physical nature. We find that the role of routinization is more nuanced than for prior technologies such as computers or robots: Although highly routinized tasks are most likely to be automated, we also observe some automation in tasks that are not routinized at the level of individuals because AI can identify patterns across individuals. As predicted, we find that automation is highest in cognitive tasks, whereas augmentation tends to be more common in hybrid tasks that require both cognitive and physical activities.

[1] Another term for such projects is "citizen science" (see Franzoni et al., 2022).

Second, we highlight that organizations may differ in whether and how they use AI even for similar tasks because they pursue different goals. While some may seek to maximize efficiency gains through automation, others may also pursue human-centered goals that require continued human involvement in tasks. In our context of crowd science, such goals include using active involvement to educate participants about particular scientific problems, training them in using scientific methods, or reducing public skepticism against science (Sauermann et al., 2020; Trouille et al., 2019). Our data show that organizations pursuing human-centered goals are less likely to use AI, which reflects lower levels of automation, not lower levels of augmentation. Moreover, organizations pursuing human-centered goals appear to design tasks differently, suggesting that task characteristics such as routinization may mediate a more fundamental role of human-centered goals. Although the particular reasons why human involvement is seen as important will differ across contexts, our results provide initial evidence that the diffusion of AI is not automatic but also shaped by the goals of organizations and their decision-makers. In other words, whether society moves towards "bad AI" that automates away human workers or "good AI" that supports workers (Acemoglu, 2023) is partly a matter of human choice. This may seem obvious to scholars who study managerial agency and strategic choice, yet it is ignored in much of the literature on the impact of AI on labor.[2]

In the next section, we develop hypotheses regarding the role of the two task characteristics that are most salient in prior literature: the degree of routinization and the cognitive vs. physical nature of tasks (Autor et al., 2003; Felten et al., 2021). We also conceptualize human-centered goals and discuss how they may shape decisions regarding whether and how to use AI. We

[2] Prior work discusses various other social, organizational, and behavioral influences on technology adoption in general (Greenwood et al., 2019; Plouffe et al., 2001). We see potential for greater integration of this literature in research on AI, but as far as we can tell, even that literature pays little attention to goals associated with human involvement per se.

then provide background on our empirical context and discuss measures, followed by the main analysis as well as supplementary analyses and robustness checks. We then discuss our contributions to the literature, including implications for the growing body of work studying what happens when AI is implemented in specific tasks such playing chess, examining patents, responding to IT requests, or selling credit cards (Allen & Choudhury, 2022; Choudhury et al., 2020; Jia et al., 2023; Krakowski et al., 2023; Lebovitz et al., 2022; Peng et al., 2023). We conclude by acknowledging limitations and opportunities for future research.

2 Conceptual background

2.1 AI and AI exposure

Scholars employ different definitions of AI (Csaszar & Steinberger, 2022) but typically refer to algorithms that use prediction to exhibit intelligent behaviors (Agrawal et al., 2018; Brynjolfsson et al., 2023; Raisch & Krakowski, 2021). While traditional rule-based algorithms are static and difficult to scale to more complex problems, AI algorithms can learn from data, adapt to new information, and handle more complex problems. Given that different AI technologies have different capabilities and requirements, we narrow our attention in two ways. First, 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. We do not consider "generative" AI that creates new artifacts based on a general underlying model (e.g., ChatGPT).[3] Second, we focus on AI as an information processing technology, distinguishing it from intelligent robots that perform physical work (Eloundou et al., 2024; Raj & Seamans, 2019).

[3] We did not find project-level applications of generative AI in our sample, suggesting that our focus facilitates theorizing without constraining the empirical analysis. Of course, future research should also study the adoption of generative AI.

When implemented in organizations, AI can have significant impacts on humans and their work. Recent studies have estimated the exposure that different occupations have to AI, revealing significant heterogeneity that likely reflects differences in underlying task characteristics (Eloundou et al., 2024; Felten et al., 2021; Sytsma & Sousa, 2023). What this exposure implies for workers, however, depends on whether organizations use AI to replace humans ("automation") or to collaborate with humans ("augmentation") (Agrawal et al., 2023; Brynjolfsson, 2022; Raisch & Krakowski, 2021). Given that automation and augmentation are not always defined consistently, we clarify our conceptualization in the next section.

2.2 Augmentation versus automation

We follow prior literature on the impacts of AI by focusing on a particular *task* within a broader job as the unit of analysis (Acemoglu & Restrepo, 2018; Autor & Handel, 2013). The job of a biological researcher at a university, for example, might involve tasks such as collecting data on different bird species and analyzing these data using statistical methods. These tasks can be further decomposed, e.g., collecting data on birds may require going out into the field to observe birds as well as determining the species of birds observed. Once we "fix" the task at a particular level of aggregation, lower-level tasks can be defined as *sub-tasks* (Agrawal et al., 2023; Puranam et al., 2014; Thompson, 1967).

Most tasks involve 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. The biological researcher, for example, could observe birds in different locations at different times of the day; recording an observation in a particular location at a particular time is one task instance.

Using these conceptual building blocks, we define *augmentation* as the collaboration of humans and AI in performing the same task instances (Lebovitz et al., 2022; Raisch & Krakowski, 2021). One common case of augmentation is that AI performs a particular task

instance and passes the result as a recommendation to the human. The human then works on the same task instance and considers the AI's recommendation to generate the final output for this instance. The task of a radiologist, for example, may be to decide whether images show cancer, and each individual image can be considered one task instance (Lebovitz et al., 2022). An AI may first analyze a particular image k and make a recommendation to the radiologist, who may then incorporate the recommendation in her own assessment (Figure 1).[4]

Another common case of augmentation is that humans and AI perform different sub-tasks required to perform the same task instance. In the field experiment described by Jia et al. (2023), for example, the task of sales representatives is to sell credit cards to potential customers. Performing this task entails two sequential sub-tasks: Identifying a potential customer (sales lead) and then persuading that potential customer to make a purchase. The organization in Jia et al. (2023) introduced AI to perform the sub-task of sales lead generation, allowing human agents to focus on the sub-task of persuading the potential customer to make a purchase.

We define *automation* as cases where AI completes some or all task instances without involvement of the human (Agrawal et al., 2023). One common case is that humans generate training data for AI by performing the first 1 to k task instances, enabling the AI to then complete remaining instances without human involvement (Figure 2). In the astrophysics project *Galaxy Zoo*, for example, humans classify thousands of images of galaxies, which are then used to train an algorithm that automates the classification of the remaining images

[4] Our definitions focus on the activities performed by humans and/or AI, regardless of performance implications. This approach facilitates measurement but also enables future research to study empirically (rather than assuming) whether augmentation makes workers more productive or whether automation makes organizations more efficient (Dell'Acqua, 2022).

(Trouille et al., 2019). Note that both humans and AI are involved in the overall task (e.g., classifying galaxies), but they perform different task instances.

A second case of automation is that AI performs all task instances, without the involvement of humans in the organization in the task. This may be accomplished by algorithms that are self-supervised or by off-the-shelf algorithms trained on external data. For example, a hospital could in principle use such algorithms to classify medical images without asking its own radiologists to create training data or to check the classifications that AI has made (Liu et al., 2024). Table 1 summarizes our conceptualization of augmentation and automation.

This discussion raises four general points. First, drawing the distinction between augmentation and automation requires "fixing" the task at a particular level of aggregation because shifting to a more disaggregated task definition may imply a re-classification from augmentation to automation. For example, if Jia et al. (2023) defined the task in their study not as "selling credit cards" but as "identifying sales leads", then their case would be classified as automation. This does not mean that the distinction between automation and augmentation is arbitrary: It means that we need to carefully choose the most meaningful definition of what is a "task" and then apply it consistently when measuring task characteristics and AI use (Felten et al., 2021; Raisch & Krakowski, 2021; Restrepo, 2023).[5]

Second, the four general cases of augmentation and automation we discussed cover many of the examples we have seen in the literature and in our data, but the framework could also be

[5] Instead of fixing a particular level of aggregation, one could try to break down tasks to the lowest possible "task primitives" but this is impractical (e.g., even "dialing a customer's phone number" could be broken down into smaller elements). More importantly, such primitives are typically not the units that workers or organizers are thinking about when performing their work or making decisions (Jia et al., 2023; Lebovitz et al., 2022). In the empirical part, we use task definitions that are meaningful in the context of the projects we study and that allow consistent measurement of task characteristics and AI use.

used to classify other workflows between humans and AI. For example, Choudhary et al. (2025) discuss "ensembling", where humans and AI perform the same task instances in parallel, with their respective outputs being aggregated through mechanisms such as averaging. Consistent with the intuition of Choudhary et al. (2025), we would classify this as augmentation because humans and AI work on the same task instances.

Third, automation and augmentation do not have to be 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. Instances classified with low confidence could be passed on to humans, along with the algorithmic classification as a recommendation. This organization would use both automation (some task instances are only performed by AI) and augmentation (some task instances are performed by both humans and AI) for the task of image classification.

Finally, much of the literature on AI exposure focuses on task characteristics without considering the particular organizational context. This is useful to predict general trends (e.g., which occupations are at greater risk of automation), but it ignores that organizations may make different choices regarding whether and how to use AI even for the same task. For example, the call center studied in Brynjolfsson et al. (2023) made an explicit decision to use AI to augment employees performing customer service, even though some other organizations automate this task. This suggests the need to study the augmentation versus automation of a particular task in the context of a particular organization. It also suggests that task characteristics alone may not be sufficient to explain whether and how organizations deploy AI. In the following, we develop our hypotheses regarding the role of both task characteristics and organizations' human-centered goals.

--- Table 1 and Figures 1-2 about here ---

2.3 The role of task characteristics

Much of the work studying automation and augmentation focuses on the mapping between task characteristics and the capabilities of a new technology to better understand which tasks can be performed by the technology. A common premise is that tasks will be automated if technological capabilities match task requirements, while augmentation may be used if tasks involve aspects that the technology cannot perform, requiring continued involvement of humans (Autor et al., 2003; Choudhury et al., 2020; Huang & Rust, 2018).[6]

We focus on two task characteristics that are particularly prominent in the prior literature and that are also important in our empirical context: Task routinization and the physical versus cognitive nature of tasks (Autor, 2015; Autor et al., 2003; Dixon et al., 2021).

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 (Autor, 2015; Pentland & Rueter, 1994). Prior technologies such as robots and computers are more likely to automate routinized vs. non-routinized tasks because these technologies typically require explicit instructions and face difficulties adjusting to changing environments. Predictions are less clear in the case of AI. On the one hand, codification may be less important if algorithms can replicate human performance without receiving explicit instructions (Aghion et al., 2018; Brynjolfsson & Mitchell, 2017). For example, an AI can identify patterns in data on past human decisions and then make similar decisions without

[6] In addition to the capabilities of a new technology, organizations will also consider its costs (Arntz et al., 2016; Autor et al., 2003). For simplicity, we assume that if a particular task characteristic makes the task easier for AI to perform, then this can translate into higher performance, lower costs, or both, either way increasing a relative advantage over humans.

explicit instructions on how to evaluate decision options. AI that receives real-time information may also be better able to adjust to changing environments than prior technologies. On the other hand, it should still be easier for AI to learn routinized tasks, e.g., because relevant variables are more likely to be codified and available in digital form and because the repetitive nature of such tasks yields a greater volume of data (Choudhury et al., 2020; Jia et al., 2023). Thus, even though AI may be able to perform less routinized tasks in principle, its relative advantages over humans are likely greater in routinized tasks for the reasons expressed in the literature on automation using prior technologies.

Hypothesis 1a: AI-based automation is more likely in tasks with high routinization than tasks with low or medium routinization.

Task routinization also improves opportunities to use AI to augment human workers. For example, an AI that provides decision support likely makes better recommendations for tasks that are routinized and for which larger volumes of data are available. However, the degree of routinization required for effective augmentation should be lower than that required for automation given that humans remain in the loop and can compensate for any deficiencies of AI. 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 receive algorithmic recommendations for relevant prior work but can use their domain-specific expertise to account for dynamic 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, efficiency-oriented organizations should automate when tasks are sufficiently routinized and should augment if tasks are not sufficiently routinized to be automated but sufficiently routinized for AI to support humans (Huang & Rust, 2018). Of

course, there may be tasks where AI does not offer significant benefits in either automation or augmentation, and these tasks will be performed solely by humans.[7]

Hypothesis 1b: AI-based augmentation is more likely in tasks with medium routinization than tasks with low or high routinization.

2.3.2 Physical vs. cognitive vs. hybrid tasks

The second main characteristic emphasized in the literature on automation is the physical vs. cognitive nature of tasks (Autor et al., 2003; Felten et al., 2021). Physical aspects of tasks involve movements and the use of physical power and skills, while cognitive aspects involve information processing and require cognitive effort. Some tasks can be classified as primarily physical or cognitive, such as lifting heavy boxes or writing a poem (Autor et al., 2003). Other tasks involve a more even balance. Collecting data on birds in nature, for example, requires physical movement to different geographical locations but also knowledge and cognitive effort to identify the species of birds that are observed. We will call tasks that involve important physical and cognitive elements "hybrid" tasks, while recognizing that the boundaries between physical, hybrid, and cognitive can be fuzzy.[8]

The core strength of AI is the processing of information, while AI – as distinct from integrated robotic systems – is unable to perform activities that require mobility or physical

[7] Our conceptual discussion and empirical analysis focus on the role of AI and humans at a given point in time. Of course, organizations' decisions to deploy AI for a given task may change as AI capabilities improve (Huang & Rust, 2018; Raisch & Krakowski, 2021). We discuss such dynamics as an important area for future research in the concluding section.

[8] Although economists and management scholars focus on the dichotomy between physical and cognitive tasks, hybrid or "dual" tasks are well-researched in physiology and medicine (Lauenroth et al., 2016). Classifying tasks again requires a fixed level of aggregation: If we decomposed the hybrid task of "collecting data on birds in nature" into "walking to find birds" and "identifying the species of birds", then the former task would be physical while the latter would be cognitive. This is not a problem if we use the same level of aggregation when measuring task characteristics and AI use (see section 2.2).

manipulation (Eloundou et al., 2024; Tschang & Almirall, 2021). As such, automation using AI should be most likely in cognitive tasks:

Hypothesis 2a: AI-based automation is more likely in cognitive than in physical or hybrid tasks.

Although AI cannot perform physical tasks, it can support humans in hybrid tasks by performing the cognitive aspects of such tasks. Using the example of collecting data on birds, an AI cannot move to different physical locations in nature but once a human researcher has made the physical move, AI could support her in identifying the species of bird that is observed. Of course, AI could 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 increasing information processing capabilities of AI, however, we expect that augmentation is more likely in hybrid tasks than in cognitive tasks because humans remain more critical in the former than the latter.

Hypothesis 2b: AI-based augmentation is more likely in hybrid tasks than physical or cognitive tasks.

2.4 The role of human-centered goals

Up to this point, we followed the common logic that the automation or augmentation of a particular task are driven by the match between task characteristics and technological capabilities (Aghion et al., 2018; Bessen, 2016; Brynjolfsson & Mitchell, 2017). Yet, automation and augmentation do not happen automatically: They reflect explicit choices by organizational decision-makers on whether and how to integrate AI into workflows. These decision-makers may seek to increase efficiency by lowering labor costs or increasing quality and speed (Autor et al., 2003; Furman & Teodoridis, 2020), but they may also pursue what we

refer to broadly as "human-centered goals": Goals that directly or indirectly require the active involvement of humans in a task.

In some cases, such a human-centered goal may be human involvement itself. For example, managers may care about providing jobs and income for workers, leading them to invest less in automation (Keum, 2023). Similarly, organizations may face pressures from unions and policy makers to preserve employment, leading them to employ more people than technically necessary (Gu et al., 2020).[9] In other cases, organizations may not care about providing employment per se, but involving people enables them to accomplish other goals such as providing workers with opportunities to develop their human capital or strengthening ties with local communities (Marquis & Battilana, 2009; Shibayama et al., 2015).[10]

Building on the prior sections, we expect that organizations pursuing only efficiency goals should automate where possible in order to increase the speed and quality of production while reducing labor costs. In contrast, organizations that also pursue human-centered goals should automate less because they seek to preserve some of the other benefits of human involvement. Although organizations could move workers threatened by automation to different and even completely new tasks, such adjustment processes take time, involve considerable costs, and tend to occur at the level of the economy rather than individual organizations (Autor, 2015). Thus, we would still expect to see less automation in organizations that pursue human-centered

[9] Managers in extremely competitive industries may not have discretion to pursue goals other than efficiency. In many industries, however, firms do have slack resources and face limited competition, allowing for managerial choice (Gu et al., 2020; Keum, 2023). See also the broader literature on how firms resolve trade-offs between efficiency and competing goals such as social responsibility or sustainability (Ethiraj & Levinthal, 2009; Stevens et al., 2015).

[10] Human-centered goals could still align with "efficiency" if performance is defined broadly (e.g., PhDs produced is an important metric for university research labs). Consistent with prior literature applying the task model to understand automation, we define performance more narrowly as production of the primary task output.

goals. On the other hand, human-centered goals should not deter organizations from using AI for augmentation because augmentation keeps humans involved, preserving employment and its associated benefits.[11]

Hypothesis 3: AI-based automation is less likely in organizations with human-centered goals than in organizations without human-centered goals.

3 Empirical context, data, and measures

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 are meaningful organizational units with their own structures and processes and can involve thousands of participants. Projects are typically led by professional scientists, including at top tier institutions such as Cornell, Oxford, and Stanford University. 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 (e.g., hobbyists in ornithology or astronomy, patients). Projects involve participants primarily in data collection and processing but also a range of other activities such as data analysis or problem solving (Haklay et al., 2021; Hecker et al., 2018).

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. For example, the main task for participants in the project *eBird* is to observe birds in

[11] We abstract from the possibility that augmentation may also affect some human-centered outcomes such as employee learning; the direction of such effects likely depends on the particular outcome and how exactly AI is implemented. Although we do not hypothesize an effect of human-centered goals on augmentation, we will probe this relationship empirically.

nature and submit records of the species and their numbers (Appendix A lists weblinks to all projects written in italics). 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 using a graphical online interface. Although some crowd members may also be engaged in secondary tasks such as providing funding or analyzing crowd-generated data, the clear task structure allows us to measure both task characteristics and AI deployment focusing on one primary task.

Second, there is significant heterogeneity across projects with respect to task characteristics and the relevance of human-centered goals. Variation in task characteristics results partly from the diversity of scientific activities performed by crowd members (e.g., classifying images, folding proteins, collecting soil samples). Organizers may also decide to design tasks differently for similar types of activities. For example, one project may ask participants to bring specimens found in nature to a physical lab (physical task) while another may ask participants to report the specimens using a digital interface (hybrid task). With respect to goals, all crowd science projects seek to generate scientific outputs such as large data sets or novel solutions to scientific problems. Some projects, however, also pursue additional goals tied to the direct involvement of crowd members in research, such as educating participants about science or building awareness of biodiversity loss. Project goals and task design are 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 goals and organizational decisions through mechanisms such as a player council in *Eterna* or discussion boards. The heterogeneity in both task characteristics and human-centered goals allows us to exploit cross-project variation to test our predictions.

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). We will discuss the generalizability of our results beyond the context of crowd science in the final section.

3.2 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 catalogue 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 tasks 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 the platform 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 AI use and other key variables. Another co-author helped resolve unclear cases throughout the coding process.[12]

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 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.

3.3 Measures

Table 2 shows summary statistics and Table 3 shows correlations.

[12] 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.

3.3.1 Independent variables

The coders carefully read project descriptions and other sources to identify the primary task performed in each project. 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 images 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 AI use) focus on the primary task only.

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*.

Physical vs. cognitive vs. hybrid tasks. Coders classified tasks as primarily cognitive if they are performed in an online environment. The rationale is that cognitive tasks are those that

involve information processing (Felten et al., 2021), and information processing is typically the focus of tasks that are performed in an online environment. Examples include the classification of digital images in *Galaxy Zoo*, the design of new RNA structures using computer-based interfaces in *Eterna*, or the tracing of fishing activity on digital maps in *Global Fishing Watch*. Tasks are classified as primarily physical when they are performed offline, without significant use of computers or smartphones. Examples include the collection of soil samples, microplastics, or animals that are then sent to principal investigators for identification and analysis (e.g., *LA Spider Survey*). We recognize that the mapping between the cognitive vs. physical nature of tasks 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 relatively more important than physical elements in online tasks, while physical elements are relatively more important than cognitive elements in offline tasks. Tasks are classified as hybrid if they involve both significant offline and online elements. Common examples include tasks to observe plants and animals in nature and submit identified records using online interfaces (e.g., *eBird*, *Monarch Larva Monitoring Project*).

Human-centered goals. To assess whether projects pursue not only scientific goals but also goals that require the active involvement of human participants, the coders reviewed sources such as project websites, blogs, as well as reports and publications. These sources sometimes included direct goal and mission statements, but the coders also looked for indirect evidence. One relevant human-centered goal is to educate participants about science. The project *DeVOTE*, for example, 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".[13] The *Butler County Stream Team* does not explicitly state educational goals but offers various courses and workshops for participants to learn about water quality as well as measures to improve water quality, strongly suggesting the presence of educational goals.[14] Other projects seek to sensitize participants to particular problems or to encourage their appreciation of nature. The *Rescue A Reef* project, for example, allows crowd members to participate in field 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."[15] Finally, some projects seek to democratize science by enabling non-professional participants to contribute to research. As stated by *Earthwise Aware*: "What sets EwA apart is its commitment to the true democratization of science… EwA breaks down traditional barriers, empowering individuals from diverse backgrounds to contribute meaningfully to ecological knowledge".[16] We argue that these and similar goals require the continued involvement of human participants – projects could not achieve these goals if they replaced participants with AI. Of course, one may be concerned that statements such as those mentioned above do not capture true goals but are marketing efforts by projects that seek to appear in a positive light. Although we cannot rule out this possibility, only a minority of projects state human-centered goals, suggesting that any potential marketing benefits are not widely expected. Moreover, if some projects only pretend to pursue human-centered goals, then the relationship between this measure and automation decisions should be attenuated, resulting in a more conservative test of our prediction.

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

[14] https://scistarter.org/butler-county-stream-team

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

[16] https://www.earthwiseaware.org/citizen-science/

3.3.2 Dependent variables

The coders searched all available sources for evidence that a project uses AI in performing 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 role. 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).[17] Note that 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 whether and how to deploy AI at the project level.

Automation. The coders assessed whether a project uses AI to perform task instances without the direct involvement of a human, which could entail different specific workflows (see section 2.2). In *Galaxy Zoo*, for example, participants classify images of galaxies and these data are then used to train an AI that classifies the remaining images in a set (Trouille et al., 2019). In *Gravity Spy*, an AI pre-classifies all images and only those that are classified with a low level of confidence are passed on to experienced crowd members to look for possible new glitch classes (Zevin et al., 2024). In *Quantum Minds*, participants apply their problem-solving skills and intuition to overcome bottlenecks in quantum physics, and the researchers then use insights into human strategies to improve problem-solving algorithms. Note that we did not track the evolution of projects to see whether and how AI has taken over task instances over

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

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",[18] 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."[19] In our sample, 7.72% of projects use AI for automation (dummy coded=1).

Augmentation. The coders assessed whether AI and humans collaborate on the same task instances. In *eBird*, for example, the AI tool Merlin helps participants classify a given bird faster and with greater accuracy. 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: "By taking data which has already been aligned by a heuristic algorithm, we allow the user to optimize where the algorithm may have failed."[20] In our sample, 8.80% of projects use AI for augmentation (dummy coded=1).[21]

3.3.3 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

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

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

[20] https://phylo.cs.mcgill.ca/.

[21] Four projects use AI to both augment and automate. The number of such projects is too small to analyze them separately.

relevant in our context, platforms may provide access to shared AI tools, reducing projects' costs of implementing AI.[22] 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 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 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 few participants and may not always be visible to external observers. Although such tasks are not considered in our coding of task characteristics or AI use, we include a dummy variable that takes the value of 1 if we find evidence of any secondary tasks.

Geographic scope. Projects differ in the geographic scope of their work. 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 level. Geographic scope may affect recruiting 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 Florida), which may have implications for the benefits and costs of using AI. We received information on geographic

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

scope directly from SciStarter and we coded it manually for Zooniverse projects (local/regional projects versus national/international projects). We create a separate dummy for projects where geographic scope could not be determined reliably.

Main research field. There may be systematic field differences in the benefits and costs of implementing AI. To control for such differences, we code projects' primary research field based on project profiles on SciStarter or Zooniverse. The large share of projects in biology as well as the earth and environmental sciences (see Table 2) likely 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 establishing the representativeness of our sample more formally.

--- Tables 2 and 3 about here ---

4 Results

4.1 Mean comparisons

Figure 3 visualizes the share of projects using augmentation and automation by the degree of routinization (Panel 1), the physical vs. cognitive nature of tasks (Panel 2), and the presence of human-centered goals (Panel 3). We will use regression analyses below to test the emerging differences statistically while accounting for controls and correlations between independent variables.

Panel 1 in Figure 3 shows that automation is most likely in highly routinized tasks, consistent with Hypothesis 1a. Interestingly, some projects also use AI to automate low routinization tasks. A closer look at projects in this category shows that several involve crowd members in experimentation or problem solving and then analyze human strategies in order to develop better algorithms. In other words, even though problem solving is low-routine and not

codified at the level of individual participants, AI can look across many individuals and learn from their collective problem-solving behaviors. Once problem-solving algorithms have improved sufficiently, they can take over remaining task instances.[23]

Augmentation is most likely in tasks with medium routinization, consistent with H1b. An example is *Frog ID*, which provides participants with AI support in the task of recording frog observations. Although some projects also employ AI to augment workers in high routinization tasks (e.g., an AI in *Living with Machines* highlights potentially relevant words in old newspaper articles to facilitate transcription), this appears less common, perhaps because automation is often a better option for such tasks.

Panel 2 shows that automation is most likely in cognitive tasks, consistent with H2a. 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 perform physical activities on its own. Augmentation is most likely in hybrid tasks, consistent with the conjecture that such tasks can benefit the most from complementary capabilities of humans and AI (H2b). A prominent example is *eBird*, which asks participants to make observations in different physical locations but facilitates the classification and recording of those observations using an AI tool.

Panel 3 shows that automation is less likely in projects with human-centered goals than in projects without human-centered goals, consistent with H3. 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 may partly reflect that this project also pursues educational goals and seeks to engage participants in environmental stewardship by involving them actively in research.[24] In contrast, Panel 3 shows no difference with respect

[23] E.g., see https://fold.it/portal/node/2008706.

[24] https://www.mysticherring.org/

to the use of AI for augmentation, suggesting that there is no inherent conflict between human-centered goals and augmentation.

4.2 Main regressions

We estimate a series of regressions with automation and augmentation as dependent variables, using linear probability models with robust standard errors (Angrist & Pischke, 2008).

Models 1-4 in Table 4 focus on task characteristics individually, showing patterns that are largely consistent with Figure 3. However, task routinization and the physical vs. cognitive nature of tasks 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 H1a (highest automation in highly routinized tasks) and H2a (highest automation in cognitive tasks). Model 6 provides only partial support for H1b (highest augmentation in medium-routinization tasks) in that the difference between low and medium routinization is not statistically significant. However, we continue to find that augmentation is more likely in medium than in high routinization tasks. Similarly, support for H2b (highest augmentation in hybrid tasks) is partial in that the difference between hybrid and cognitive tasks is not significant, but augmentation remains significantly more likely in hybrid than physical tasks.

Model 7 shows that projects with human-centered goals are less likely to use automation than projects without, supporting H3. In contrast, Model 8 shows no association between human-centered goals and the use of AI to augment human participants.

Models 9 and 10 include all independent variables simultaneously. The coefficients of task characteristics are virtually unchanged compared to models 5 and 6. The positive coefficient of human-centered goals in the regression of automation, however, becomes insignificant. This likely reflects that projects with human-centered goals tend to design tasks that are less routinized and less likely to be purely cognitive. We will further explore these relationships in supplementary analyses.

--- Table 4 about here ---

4.3 Supplementary analyses and robustness checks

We perform several supplementary analyses and robustness checks. First, we use ordered logistic regressions to regress task characteristics on human-centered goals (Table 5). We find that tasks in projects with human-centered goals tend to be less routinized (model 1) and less likely to be purely cognitive (model 2). 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 other goals such as education or awareness building. More realistically, however, task characteristics are not exogenous and partly reflect organizers' goals. Organizers who value human engagement or seek to 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 human-centered goals on automation. Future research is required to investigate the causal nature of these relationships in more detail.

Second, we collapse the measures of automation and augmentation into the single measure *AnyAI*, which we interpret as a proxy for the (realized) 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* (Table 5, model 3), even though it predicted both augmentation and automation in the main regressions (Table 4). The reason is that high levels of routinization predict higher levels of automation but lower levels of augmentation. Averaging these contrasting relationships by using the aggregate measure of AI exposure obscures important differences. We find similar patterns when looking at the cognitive vs. physical nature of tasks and the presence of human-centered goals (models 3-5): The coefficients are notably smaller when using the *AnyAI* measure, obscuring the significant – but different – relationships that these variables have with both augmentation and automation.

Third, we explore whether human-centered goals more strongly predict whether and how AI is used if tasks are more suitable for AI in terms of routinization or their cognitive vs. physical nature. Models 6-7 include interaction terms between human-centered goals and task characteristics; none of the interaction terms is statistically significant.

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 use of AI if larger projects have greater resources or expertise to develop AI technologies, have more work to be performed, or generate more data that AI can learn from. On the other hand, AI use 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 code 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 create an additional dummy equal to one if size information is missing. Models 1-6 reported in Table 6 show that project size has a positive relationship with automation, suggesting that positive effects of size on AI use (possibly via greater resources or

data availability) outweigh potential negative effects of automation on size. Size also has the expected positive relationship with augmentation. More importantly, the coefficients of task characteristics and human-centered goals are very similar to those in the main analyses, suggesting that our results are robust to potential confounding effects of project size.

--- Tables 5 and 6 about here ---

5 Discussion

5.1 Summary

We used data from more than 1,200 crowd science projects to explore how the deployment of AI to automate or augment human workers relates to task characteristics as well as organizational goals. We find that automation is most likely in projects involving highly routinized tasks and in projects involving cognitive (vs. physical) activities. In contrast, augmentation is most likely in tasks with medium routinization and in "hybrid" tasks that involve both cognitive and physical elements. These results are consistent with the idea of a progression from no use of AI, to augmentation if AI offers some benefits but cannot fully replace humans, to automation if task requirements match AI capabilities (Huang & Rust, 2018).

Our results also suggest that whether and how AI is deployed depends on organizational goals, a factor that has been largely ignored in prior literature on the impacts of AI. More specifically, projects that pursue not only efficiency but also goals that require active human involvement (e.g., education or building awareness among participants) are less likely to use automation than other projects but they are equally likely to use AI to augment. This relationship may partly be mediated by task characteristics because organizations with human-centered goals are less likely to design highly routinized and purely cognitive tasks.

5.2 Limitations

We recognize two important limitations. First, our coders performed extensive searches to see whether and how projects are using AI, but they 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. Future research should replicate our results using different approaches to measure AI. Relatedly, all crowd science projects by definition involve humans in the primary task and there are no projects where the primary task is fully automated (fourth case discussed in section 2.2). As such, it is not clear whether and how our results apply to this extreme end of the spectrum.

Second, our data are cross-sectional, yielding correlational but not causal evidence. In particular, we cannot clearly identify the extent to which task characteristics are determined by the intrinsic nature of certain research activities and to what extent organizers design tasks differently because they pursue human-centered goals (Hackman & Oldham, 1976; National Academies, 2018). Task characteristics would be important either way, but the second case would imply a more fundamental role of organizational goals. Relatedly, we control for several potentially confounding 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. Future research could complement our correlational evidence with a greater focus on identification, e.g., by tracking projects over time, exploiting shocks (such as the arrival of new pre-packaged AI technologies or legal worker protections), or using lab experiments to study organizers' adoption decisions.

5.3 Generalizability beyond crowd science

Crowd science has developed rapidly over the past decades, involving millions of participants and producing publications in top tier journals such as *Nature* and *PNAS* (Franzoni

et al., 2022). The project *eBird* has contributed roughly half of the data in the Global Biodiversity Information Facility, and *Foldit* was an important step in David Baker's path towards the 2024 Nobel Prize in Chemistry.[25] Although this makes crowd science a relevant context in itself, we now consider whether the results may generalize to other contexts.

A first relevant consideration is that many CS projects involve small tasks of low complexity, which partly reflects that such tasks are more suitable for parallel processing by a distributed workforce (Boudreau & Lakhani, 2013). The range of tasks in our sample is much broader than on "micro-task" platforms such as MTurk or Clickworker, however, in that some tasks involve complex problem solving and many involve physical activities. Although we did not observe extremely complex tasks such as running a clinical trial or writing a full research paper, even traditional organizations typically decompose such large tasks before assigning them to individual workers or to AI (Brynjolfsson et al., 2023; Jia et al., 2023; Shibayama et al., 2015). Thus, the tasks we study may not differ that much from the tasks typically performed by a significant share of the general workforce. More importantly, our sample shows sufficient variation in both task routinization and the cognitive vs. physical nature of tasks, and we expect that including an even broader range of tasks would only strengthen the observed relationships between task characteristics and the (non-)use of AI.

Second, participants in crowd science tend to be quite diverse with respect to their demographic characteristics, skills, and physical location (Cooper et al., 2021). Participation by individual crowd members also tends to be short-lived such that organizations need to cope with high turnover (Sauermann & Franzoni, 2015). These features are not unique per se, given the increasing diversity in traditional organizations as well as high rates of turnover in many

[25] https://ebird.org/news/2021-year-in- review; https://www.nobelprize.org/uploads/2024/12/baker-lecture.pdf

sectors (U.S. Bureau of Labor Statistics, 2024). Still, they may affect both the incentives to implement AI (e.g., to address labor shortages) and the feasibility of doing so (e.g., training workers to collaborate with AI). This, in turn, may affect the observed rates of AI use and even the relative advantages of automation versus augmentation. Although these factors should not fundamentally change the relationships we study, this discussion suggests opportunities for future work to explore the relationship between AI adoption and workforce characteristics.

Third, the specific human-centered goals we observe are likely context specific (e.g., educating participants about science). Yet, the broader construct is relevant across organizational contexts: Some for-profit organizations and especially social enterprises also employ workers for reasons such as helping them build human capital or because they see providing jobs as a service to their communities (Battilana & Lee, 2014; Keum & Meier, 2023). Relatedly, although one may expect that human-centered goals are more common in crowd science than in other contexts, projects with such goals are a minority in our sample and their share is not obviously higher than that of business managers expressing pro-social goals or of companies who state objectives such as job creation or employee development (Keum, 2023).

Finally, crowd science participants are unpaid. Unpaid labor is important also in the non-profit sector and in social enterprises, and even many for-profit firms enjoy wage discounts in return for offering interesting work or a mission that is aligned with workers' preferences (Bode & Singh, 2018; Stern, 2004). Nevertheless, this feature is important because "free" labor may reduce the incentives for organizations to implement AI, especially for the purpose of reducing labor costs. In some sense, being able to abstract from direct labor costs is an advantage of our empirical context because our theorizing focused on the alignment between task characteristics and AI capabilities rather than on labor costs, consistent with much of the prior literature. Even without direct labor costs, crowd science projects still 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 augments crowd members (Ponti & Seredko, 2022; Trouille et al., 2019). Although we believe that our key results generalize to contexts where workers are paid, future research should study more explicitly the cost side of decisions to use AI, including the costs of deploying AI as well as the costs of involving humans (see Acemoglu & Restrepo, 2018; Svanberg et al., 2024).

5.4 Contributions

Our study contributes to a growing body of research on AI and its implications for human workers. First, we complement studies on the AI exposure of different tasks or occupations (Eloundou et al., 2024; Felten et al., 2021; Svanberg et al., 2024) by moving from ratings of potential exposure to realized exposure in a particular context. Moreover, we show that exposure in the form of automation and in the form of augmentation have very different relationships with task characteristics, highlighting the value of distinguishing automation and augmentation in future work (Agrawal et al., 2023; Raisch & Krakowski, 2021). Our conceptual discussion in section 2.2 may also help move research using this distinction towards a shared understanding of how augmentation and automation can be defined and measured. One important conceptual insight is that it is critical to fix the level of task aggregation when measuring automation, augmentation, and task characteristics. Another important insight is that augmentation should be defined as human-AI collaboration on individual *task instances* (e.g., classifying particular images) rather than on the overall *task* (e.g., classifying all images) because the latter would (mis-)classify as augmentation cases where humans create training data that are used to automate all remaining task instances.

Second, we provide novel evidence on the relationships between task characteristics and AI deployment by analyzing data on projects that each entail a well-defined primary task. Compared to studying the effects of AI on broader jobs or occupations, this approach enables

us to focus more clearly on automation versus augmentation for a given task while ruling out confounding reinstatement effects (where one task is automated away but another task emerges in the same occupation) (Autor, 2015; Felten et al., 2021). Our results regarding routinization speak to an ongoing debate regarding whether AI is similar to prior technologies such as robots and computers in that it requires routinization and codification (Choudhury et al., 2020; Jia et al., 2023) or whether AI can also learn tasks when routinization is low or humans cannot express why they made particular decisions (Aghion et al., 2018; Brynjolfsson & Mitchell, 2017). Supporting the second perspective, we identified several cases where AI is used to automate low routinization tasks. Overall, however, we find a strong positive relationship between routinization and automation, suggesting that routinized tasks are still *relatively* easier to automate and will (initially) be where most of the impact of AI-based automation is felt.

We also provide empirical evidence on the relationships between AI use and the cognitive versus physical nature of tasks. Perhaps more importantly, we move beyond the conventional dichotomy of cognitive versus physical tasks (Autor et al., 2003) to highlight hybrid tasks as both conceptually and empirically important: While the dichotomy lends itself to study automation (yes for cognitive tasks, no for physical tasks), distinguishing hybrid tasks is useful to study augmentation because these tasks are where complementarities between humans and AI are most pronounced. Of course, physical aspects of hybrid tasks could also be performed by robots rather than humans, suggesting interesting avenues for future research on the use and impact of AI-controlled robotic systems.

Third, we highlight the need to think beyond task characteristics and efficiency to consider the goals of organizations and their decision-makers, including goals that require the active involvement of humans in tasks. Although such human-centered goals have been largely ignored in prior studies on AI, they may have profound implications for how potential AI exposure (largely driven by AI capabilities and task characteristics) ultimately plays out for

workers and their jobs. For example, while a narrowly efficiency-oriented model would suggest that increasing AI capabilities will lead to increasing automation (Huang & Rust, 2018), employment concerns may lead some organizations to use increasing AI capabilities for more extensive augmentation. Indeed, human-centered goals may lead organizations to fundamentally rethink work processes to create opportunities for human-AI collaboration that end up creating more value than seemingly superior automation (Acemoglu, 2023). Of course, our results on human-centered goals raise questions for future work. Among others, we need deeper insights into mechanisms – human-centered goals may shape AI adoption not only directly but also indirectly, e.g., if they lead organizations to design tasks differently or to train their workers to better collaborate with AI. It might also be important to distinguish different types of human-centered goals and why some organizations endorse them while others do not. Indeed, AI adoption may be an interesting context to study more generally how organizations pursue multiple and potentially conflicting goals (Ethiraj & Levinthal, 2009; Stevens et al., 2015).

By studying organizations' decisions on whether and how to use AI, we also complement a fast growing body of research that examines what happens when AI is (experimentally) deployed in specific tasks such as playing chess, examining patents, responding to IT requests, selling credit cards, developing code, or interpreting medical images (Allen & Choudhury, 2022; Choudhury et al., 2020; Jia et al., 2023; Krakowski et al., 2023; Lebovitz et al., 2022; Peng et al., 2023). One implication of our results is that future work studying such specific cases should consider more explicitly the nature of the task and what it implies for the potential of AI to augment versus automate human workers. This may enable a better understanding of why we observe particular human-AI interactions in a given case, or why performance and employment effects of AI differ across cases. Future work studying cases of AI adoption in real organizations should also ask why organizations implemented AI in the first place, and

why they use it to automate versus augment their workers. Although economic logic might lead one to assume productivity-related motives, some organizations may also pursue other goals that shape whether and how AI is deployed. This, in turn, has implications for measuring and interpreting the effects of AI on both workers and organizational performance.

5.5 Additional opportunities for future research

We conclude by highlighting two particularly interesting broader areas for future work. First, research is needed on the perspective of workers who are affected by AI. Our finding that automation is most likely in routinized tasks (e.g., classifying images) suggests that workers may be able to focus on less routinized and more challenging tasks instead. Of course, workers who are unable or un-interested to move to such other tasks may have nothing left to do. Interestingly, some crowd science projects that ask contributors to generate training data for automation then encourage them to move to other projects that need human help with similar tasks (e.g., classifications of solar storms instead of galaxies, see Trouille et al. (2019)). Thus, measuring automation and augmentation at the level of individual organizations may enable scholars to study how adjustment processes play out not only by moving workers to different types of tasks (Autor, 2015) but also across organizations that differ in their use of AI for the same types of tasks. With respect to augmentation, research should study whether the different types of augmentation identified in our conceptual discussion have different implications for workers. For example, augmentation in the form of recommendations may have different implications for job satisfaction and learning opportunities than augmentation where AI takes over sub-tasks (see Hackman & Oldham, 1976). It is also important to better understand when augmentation requires additional (AI-related) skills from employees (Krakowski et al., 2023) and when it lowers skill requirements. Many of the cases we have seen seem to do the latter

(e.g., species recommendations in *eBird*), enabling participation by a broader range of people and allowing projects to operate at a larger scale.

Second, future work is needed to study the dynamics of automation and augmentation and of related organizational changes. For example, our results regarding the role of task characteristics raise the question of whether and when organizations can decompose and re-design tasks to make them more suitable for AI. At the same time, the relevance of task characteristics may decrease over time as AI improves and becomes more effective also in performing tasks that are currently difficult to augment or automate. Similarly, it will be interesting to study whether organizations that automate one task create new tasks that can keep their workers occupied – and whether this depends on the degree to which they pursue human-centered goals. Finally, organizations face great uncertainty regarding how AI capabilities will develop for a given task, how AI will interact with workers and other parts of an organization, and what impacts AI will ultimately have on organizational performance (Frank et al., 2019; Lebovitz et al., 2022). Thus, organizations will have to experiment and may change their decisions on whether and how to deploy AI (Huang & Rust, 2018; Raisch & Krakowski, 2021). Future research should study these adjustment processes and examine what patterns emerge as AI technologies mature and organizations find out how to best deploy AI to accomplish their efficiency- and human-centered goals.

References

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

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.

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.

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

Arntz, M., Gregory, T., & Zierahn, U. 2016. The risk of automation for jobs in OECD countries: A comparative analysis.

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.

Battilana, J., & Lee, M. 2014. Advancing research on hybrid organizing–Insights from the study of social enterprises. ***Academy of Management Annals***, 8(1): 397-441.

Bessen, J. E. 2016. How computer automation affects occupations: Technology, jobs, and skills. ***Boston Univ. School of Law Working Paper***.

Bianchini, S., Müller, M., & Pelletier, P. 2022. Artificial intelligence in science: An emerging general method of invention. ***Research Policy***, 51(10): 104604.

Bode, C., & Singh, J. 2018. Taking a hit to save the world? ***Strategic Management Journal***, 39(4): 1003-1030.

Boudreau, K. J., & Lakhani, K. R. 2013. Using the crowd as an innovation partner. ***Harvard Business Review***, 91(4): 60-69, 140.

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.

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.

Cooper, C. B., Hawn, C. L., Larson, L. R., Parrish, J. K., Bowser, G., et al. 2021. Inclusion in citizen science: The conundrum of rebranding. ***Science***, 372(6549): 1386-1388.

Csaszar, F. A., & Steinberger, T. 2022. Organizations as artificial intelligences: The use of artificial intelligence analogies in organization theory. ***Academy of Management Annals***, 16(1): 1-37.

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***.

Dell'Acqua, F. 2022. Falling asleep at the wheel: Human/AI Collaboration in a field experiment on HR recruiters: 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.

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.

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.

Furman, J. L., & Teodoridis, F. 2020. Automation, research technology, and researchers' trajectories: Evidence from computer science and electrical engineering. ***Organization Science***, 31(2): 330-354.

Greenwood, B. N., Agarwal, R., Agarwal, R., & Gopal, A. 2019. The role of individual and organizational expertise in the adoption of new practices. ***Organization Science***, 30(1): 191-213.

Gu, Z., Tang, S., & Wu, D. 2020. The political economy of labor employment decisions: Evidence from China. ***Management Science***, 66(10): 4703-4725.

Guenzel, M., Hamilton, C., & Malmendier, U. 2023. CEO Social Preferences and Layoffs.

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.

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.

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

Keum, D. 2023. Managerial Prosocial Preferences and Automation Innovation. ***SSRN Working Paper #4503762***.

Keum, D. D., & Meier, S. 2023. License to layoff? Unemployment insurance and the moral cost of layoffs. ***Organization Science***.

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.

Krakowski, S., Luger, J., & Raisch, S. 2023. Artificial intelligence and the changing sources of competitive advantage. ***Strategic Management Journal***, 44(6): 1425-1452.

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).

Marquis, C., & Battilana, J. 2009. Acting globally but thinking locally? The enduring influence of local communities on organizations. ***Research in Organizational Behavior***, 29: 283-302.

Mollick, E., & Mollick, L. 2023. Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts, ***Working Paper***.

National Academies. 2018. ***Learning through citizen science: Enhancing opportunities by design***: National Academies Press.

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.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. 2023. The impact of AI on developer productivity: Evidence from github copilot. ***arXiv preprint arXiv:2302.06590***.

Pentland, B. T., & Rueter, H. H. 1994. Organizational routines as grammars of action. ***Administrative Science Quarterly***: 484-510.

Plouffe, C. R., Hulland, J. S., & Vandenbosch, M. 2001. Richness versus parsimony in modeling technology adoption decisions—understanding merchant adoption of a smart card-based payment system. ***Information Systems Research***, 12(2): 208-222.

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., Alexy, O., & Reitzig, M. 2014. What's "new" about new forms of organizing? ***Academy of Management Review***, 39(2): 162-180.

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

Raj, M., & Seamans, R. 2019. Primer on artificial intelligence and robotics. ***Journal of Organization Design***, 8(1): 11.

Restrepo, P. 2023. Automation: Theory, Evidence, and Outlook, ***NBER Working Paper #31910***.

Sauermann, H., & Franzoni, C. 2015. Crowd science user contribution patterns and their implications. ***Proceedings of the National Academy of Sciences***, 112(3): 679-684.

Sauermann, H., Vohland, K., Antoniou, V., Balaz, B., Goebel, C., et al. 2020. Citizen science

and sustainability transitions. ***Research Policy***, 49(5): 103978.

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.

Stern, S. 2004. Do scientists pay to be scientists? ***Management Science***, 50(6): 835-853.

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.

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.

Thompson, J. D. 1967. ***Organizations in Action***. New York: McGraw-Hill.

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.

U.S. Bureau of Labor Statistics. 2024. Job openings and labor turnover survey (JOLTS).

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: Augmentation vs. automation

	Augmentation	Automation
Definition	AI collaborates with humans working on the *same task instances* (e.g., classifying the same image).	AI replaces humans by working on *different or even all task instances* (e.g., classifying different images; classifying all images).
Common cases and examples	• AI provides recommendations about the same task instance to the human. E.g., classification of medical images in Lebovitz et al. (2022). • AI and human work on different sub-tasks of the same task instance. E.g., generating sales lead vs. convincing the customer in Jia et al. (2023).	• Humans generate training data by working on some task instances, AI then performs remaining task instances. E.g., classifications of galaxies in Trouille et al. (2019). • AI performs all task instances without human involvement. E.g., a pre-trained algorithm can classify all medical images for a hospital (Liu et al., 2024).

Table 2: Summary statistics

Variable	N	Mean	Min	Max
Augmentation	1,205	0.09	0	1
Automation	1,205	0.08	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 nature: Physical	1,205	0.13	0	1
Task nature: Hybrid	1,205	0.68	0	1
Task nature: Cognitive	1,205	0.19	0	1
Human-centered goals	1,205	0.17	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 sciences	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

Table 3: Correlations

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Augmentation	1														
2	Automation	-0.0459	1													
3	Routinization: Low	-0.0524	-0.0318	1												
4	Routinization: Medium	0.0897*	-0.4397*	-0.5721*	1											
5	Routinization: High	-0.0650*	0.5573*	-0.1379*	-0.7335*	1										
6	Task nature: Physical	-0.1206*	-0.1123*	0.1134*	0.046	-0.1496*	1									
7	Task nature: Hybrid	0.1397*	-0.4181*	-0.0966*	0.5244*	-0.5532*	-0.5616*	1								
8	Task nature: Cognitive	-0.0625*	0.5922*	0.0176	-0.6616*	0.7844*	-0.1897*	-0.7059*	1							
9	Human-centered goals	-0.0125	-0.0956*	0.1400*	0.0179	-0.1376*	0.1769*	-0.0299	-0.1160*	1						
10	Platform affiliated	0.2386*	0.1663*	-0.0769*	-0.1609*	0.2580*	-0.1458*	-0.046	0.1794*	-0.0462	1					
11	Age	-0.0272	-0.2200*	-0.0164	0.2636*	-0.3047*	0.0703*	0.2247*	-0.3268*	0.0458	-0.2145*	1				
12	Secondary tasks	-0.0502	-0.0299	0.2285*	-0.0877*	-0.0835*	0.1792*	-0.1225*	-0.0081	0.1996*	-0.0995*	0.0199	1			
13	Geo scope: Local/regional	-0.0268	-0.1828*	-0.0045	0.2182*	-0.2597*	0.1182*	0.1565*	-0.2869*	0.0667*	-0.2397*	0.7556*	0.0334	1		
14	Geo scope: National/intern.	0.0329	0.2570*	-0.0013	-0.2944*	0.3566*	-0.1165*	-0.2466*	0.3924*	-0.0237	0.1924*	-0.6645*	-0.0108	-0.7724*	1	
15	Geo scope: Missing	-0.006	-0.0874*	0.0088	0.0866*	-0.1119*	-0.0143	0.1127*	-0.1215*	-0.0681*	0.0915*	-0.2059*	-0.0357	-0.4250*	-0.2467*	1
16	Mainfield: Biology	0.0721*	-0.1046*	-0.0326	0.1351*	-0.1362*	-0.0426	0.1897*	-0.1886*	-0.1161*	-0.0401	0.1317*	-0.0395	0.1436*	-0.1638*	0.0144
17	Mainfield: Earth/env. sci.	-0.0347	-0.1418*	-0.0915*	0.1670*	-0.1259*	0.1139*	0.0752*	-0.1867*	0.0925*	0.0051	0.0363	-0.0072	0.0097	-0.0641*	0.0766*
18	Mainfield: Physical sciences	-0.0031	0.3238*	0.0489	-0.2650*	0.2795*	-0.0763*	-0.2111*	0.3158*	-0.0453	0.1090*	-0.1588*	0.015	-0.1178*	0.1831*	-0.0813*
19	Mainfield: Social sci./human.	-0.0536	0.0321	0.1538*	-0.2126*	0.1292*	-0.0046	-0.2073*	0.2499*	0.1134*	0.0127	-0.1793*	0.0648*	-0.1210*	0.1872*	-0.0821*
20	Mainfield: Other	-0.0252	0.1584*	0.0534	-0.1516*	0.1388*	-0.0630*	-0.1383*	0.2181*	-0.0124	-0.0483	-0.014	0.0185	-0.0859*	0.1085*	-0.0236

Note: *=Significant at 5%

Table 4: Main regressions of automation and augmentation

	1	2	3	4	5	6	7	8	9	10
	LPM	LPM	LPM	LPM	LPM	LPM	LPM	LPM	LPM	LPM
Variable	Autom.	Augm.	Autom.	Augm.	Autom.	Augm.	Autom.	Augm.	Autom.	Augm.
Routinization: Low	-0.333**	-0.043+			-0.158**	-0.034			-0.157**	-0.035
	[0.045]	[0.022]			[0.055]	[0.022]			[0.055]	[0.022]
Routinization: Medium	-0.354**	omitted			-0.152**	omitted			-0.152**	omitted
	[0.040]				[0.052]				[0.052]	
Routinization: High	omitted	-0.127**			omitted	-0.119**			omitted	-0.119**
		[0.023]				[0.035]				[0.035]
Task nature: Physical			-0.353**	-0.086**	-0.251**	-0.081**			-0.251**	-0.082**
			[0.035]	[0.011]	[0.046]	[0.011]			[0.046]	[0.012]
Task nature: Hybrid			-0.368**	omitted	-0.262**	omitted			-0.262**	omitted
			[0.035]		[0.046]				[0.046]	
Task nature: Cognitive			omitted	-0.097**	omitted	-0.016			omitted	-0.015
				[0.020]		[0.031]				[0.031]
Human-centered goals							-0.051**	0.010	-0.002	0.007
							[0.015]	[0.022]	[0.012]	[0.022]
Platform affiliated	0.013	0.157**	0.030*	0.143**	0.018	0.151**	0.060**	0.141**	0.018	0.151**
	[0.013]	[0.020]	[0.013]	[0.019]	[0.012]	[0.020]	[0.015]	[0.019]	[0.012]	[0.020]
Age	-0.010	-0.006	-0.009	-0.006	-0.007	-0.008	-0.024*	-0.001	-0.007	-0.008
	[0.011]	[0.009]	[0.011]	[0.009]	[0.011]	[0.009]	[0.012]	[0.010]	[0.011]	[0.009]
Secondary tasks	0.004	-0.025	-0.019	-0.007	-0.005	-0.010	-0.013	-0.025	-0.004	-0.011
	[0.023]	[0.023]	[0.023]	[0.022]	[0.022]	[0.023]	[0.027]	[0.023]	[0.022]	[0.024]
Geo scope: Local/regional	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Geo scope: National/intern.	0.005	0.009	-0.007	0.002	-0.009	0.002	0.035	-0.001	-0.009	0.003
	[0.028]	[0.028]	[0.027]	[0.028]	[0.027]	[0.028]	[0.030]	[0.028]	[0.027]	[0.028]
Geo scope: Missing	-0.023	-0.045	-0.028	-0.046	-0.020	-0.051+	-0.067**	-0.032	-0.021	-0.050+
	[0.021]	[0.029]	[0.021]	[0.029]	[0.021]	[0.029]	[0.024]	[0.029]	[0.021]	[0.028]
Mainfield: Biology	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Mainfield: Earth/env. sci.	-0.018+	-0.041*	-0.015	-0.034*	-0.016	-0.035*	-0.019	-0.038*	-0.015	-0.036*
	[0.010]	[0.017]	[0.010]	[0.017]	[0.010]	[0.017]	[0.012]	[0.018]	[0.010]	[0.018]
Mainfield: Physical sciences	0.197**	-0.017	0.149**	-0.023	0.147**	-0.017	0.314**	-0.062	0.147**	-0.018
	[0.053]	[0.036]	[0.052]	[0.036]	[0.052]	[0.036]	[0.057]	[0.038]	[0.052]	[0.036]
Mainfield: Social sci./human.	-0.039	-0.065**	-0.119**	-0.055*	-0.101*	-0.061**	0.024	-0.093**	-0.101*	-0.063**
	[0.039]	[0.024]	[0.042]	[0.023]	[0.042]	[0.023]	[0.042]	[0.027]	[0.042]	[0.024]
Mainfield: Other	0.107*	-0.009	0.047	-0.008	0.054	-0.010	0.191**	-0.041	0.054	-0.011
	[0.050]	[0.033]	[0.046]	[0.032]	[0.046]	[0.030]	[0.052]	[0.033]	[0.046]	[0.031]
Constant	0.412**	0.103+	0.413**	0.115+	0.451**	0.125*	0.158*	0.065	0.451**	0.125*
	[0.077]	[0.058]	[0.072]	[0.059]	[0.075]	[0.059]	[0.072]	[0.060]	[0.075]	[0.058]
Observations	1,205	1,205	1,205	1,205	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.356	0.086	0.388	0.086	0.403	0.094	0.196	0.067	0.403	0.094
df	11	11	11	11	13	13	10	10	14	14

Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%. LPM = Linear Probability Model (OLS).

Table 5: Supplementary analyses

Variable	1 ologit Routinization	2 ologit CogHybPhys	3 LPM AnyAI	4 LPM AnyAI	5 LPM AnyAI	6 LPM Autom.	7 LPM Augm.
Routinization: Low			-0.093		-0.093	omitted	omitted
			[0.060]		[0.061]		
Routinization: Medium			-0.049		-0.049	0.002	0.019
			[0.056]		[0.056]	[0.025]	[0.030]
Routinization: High			omitted		omitted	0.138*	-0.098*
						[0.060]	[0.046]
Task nature: Physical			-0.291**		-0.292**	omitted	omitted
			[0.049]		[0.049]		
Task nature: Hybrid			-0.220**		-0.220**	-0.012*	0.064**
			[0.049]		[0.049]	[0.006]	[0.012]
Task nature: Cognitive			omitted		omitted	0.269**	0.051
						[0.049]	[0.034]
Human-centered goals	-0.978**	-1.066**		-0.038	0.006	-0.015	-0.082*
	[0.204]	[0.186]		[0.025]	[0.025]	[0.027]	[0.037]
Human-c. X Medium rout.						0.017	0.054
						[0.029]	[0.039]
Human-c. X High rout.						0.134	0.044
						[0.214]	[0.057]
Human-c. X Hybrid						0.011	0.060+
						[0.010]	[0.031]
Human-c. X Cognitive						-0.156	0.025
						[0.113]	[0.040]
Platform affiliated	0.935**	0.753**	0.169**	0.201**	0.169**	0.016	0.153**
	[0.160]	[0.139]	[0.023]	[0.023]	[0.023]	[0.012]	[0.020]
Age	-0.160+	-0.137	-0.016	-0.025+	-0.016	-0.008	-0.009
	[0.083]	[0.091]	[0.014]	[0.014]	[0.014]	[0.011]	[0.009]
Secondary tasks	-1.438**	-0.734*	-0.009	-0.034	-0.010	-0.005	-0.009
	[0.303]	[0.309]	[0.032]	[0.034]	[0.033]	[0.022]	[0.024]
Mainfield: Biology	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Mainfield: Earth/env. sci.	0.135	-0.352*	-0.049*	-0.056**	-0.050*	-0.016	-0.035+
	[0.134]	[0.141]	[0.020]	[0.021]	[0.020]	[0.010]	[0.018]
Mainfield: Physical sciences	1.258**	2.301**	0.110*	0.226**	0.110*	0.149**	-0.014
	[0.362]	[0.328]	[0.055]	[0.057]	[0.055]	[0.052]	[0.036]
Mainfield: Social sci./human.	0.042	1.947**	-0.153**	-0.068	-0.154**	-0.092*	-0.059*
	[0.455]	[0.367]	[0.047]	[0.047]	[0.047]	[0.043]	[0.024]
Mainfield: Other	0.940*	2.061**	0.038	0.137*	0.037	0.051	-0.011
	[0.420]	[0.328]	[0.049]	[0.055]	[0.049]	[0.046]	[0.031]
Geo scope: Local/regional	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Geo scope: National/intern.	0.641*	0.989**	-0.005	0.033	-0.005	-0.009	0.001
	[0.254]	[0.241]	[0.037]	[0.037]	[0.037]	[0.027]	[0.028]
Geo scope: Missing	-0.803**	-0.296	-0.071*	-0.097**	-0.070*	-0.019	-0.052+
	[0.246]	[0.235]	[0.034]	[0.035]	[0.034]	[0.021]	[0.029]
Constant			0.436**	0.224**	0.436**	0.050	0.041
			[0.089]	[0.083]	[0.089]	[0.071]	[0.064]
Observations	1,205	1,205	1,205	1,205	1,205	1,205	1,205
R-squared			0.223	0.164	0.223	0.406	0.096
df	10	10	13	10	14	18	18

Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%. *CogHybPhys* is the 3-point measure of task nature (1=physical, 2=hybrid, 3=cognitive). *AnyAI* takes the value of 1 if a project uses automation and/or augmentation.

Table 6: 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	-0.139*	-0.031			-0.138*	-0.031
	[0.055]	[0.021]			[0.056]	[0.021]
Routinization: Medium	-0.135**	omitted			-0.135**	omitted
	[0.052]				[0.052]	
Routinization: High	omitted	-0.154**			omitted	-0.154**
		[0.035]				[0.035]
Task nature: Physical	-0.236**	-0.069**			-0.235**	-0.069**
	[0.045]	[0.011]			[0.045]	[0.011]
Task nature: Hybrid	-0.253**	omitted			-0.252**	omitted
	[0.044]				[0.044]	
Task nature: Cognitive	omitted	-0.022			omitted	-0.022
		[0.029]				[0.029]
Human-centered goals			-0.047**	0.009	-0.005	0.001
			[0.015]	[0.022]	[0.013]	[0.022]
Size: >=1,000	omitted	omitted	omitted	omitted	omitted	omitted
Size: 1,001-10,000	0.055**	0.067*	0.134**	0.034	0.055**	0.067*
	[0.021]	[0.030]	[0.023]	[0.029]	[0.021]	[0.030]
Size: >10,000	0.072**	0.074*	0.181**	0.036	0.073**	0.074*
	[0.028]	[0.035]	[0.034]	[0.035]	[0.028]	[0.036]
Size: Missing	0.014	-0.056**	0.037**	-0.064**	0.014	-0.056**
	[0.011]	[0.018]	[0.012]	[0.018]	[0.011]	[0.018]
Platform affiliated	0.019	0.142**	0.054**	0.127**	0.019	0.142**
	[0.012]	[0.019]	[0.014]	[0.018]	[0.012]	[0.019]
Age	-0.011	-0.010	-0.032**	0.000	-0.011	-0.010
	[0.011]	[0.009]	[0.012]	[0.010]	[0.011]	[0.009]
Secondary tasks	-0.005	-0.021	-0.005	-0.031	-0.004	-0.022
	[0.022]	[0.023]	[0.026]	[0.023]	[0.022]	[0.024]
Mainfield: Biology	omitted	omitted	omitted	omitted	omitted	omitted
Mainfield: Earth/env. sci.	-0.015	-0.030+	-0.015	-0.033+	-0.014	-0.030+
	[0.010]	[0.017]	[0.012]	[0.018]	[0.010]	[0.017]
Mainfield: Physical sciences	0.139**	-0.024	0.269**	-0.072+	0.139**	-0.024
	[0.051]	[0.035]	[0.054]	[0.038]	[0.051]	[0.035]
Mainfield: Social sci./human.	-0.100*	-0.054*	0.013	-0.089**	-0.099*	-0.054*
	[0.042]	[0.023]	[0.042]	[0.027]	[0.042]	[0.024]
Mainfield: Other	0.055	-0.000	0.175**	-0.039	0.056	-0.000
	[0.045]	[0.031]	[0.050]	[0.032]	[0.045]	[0.031]
Geo scope: Local/regional	omitted	omitted	omitted	omitted	omitted	omitted
Geo scope: National/intern.	-0.011	-0.004	0.022	-0.011	-0.011	-0.004
	[0.027]	[0.028]	[0.030]	[0.028]	[0.027]	[0.028]
Geo scope: Missing	-0.019	-0.047+	-0.057*	-0.026	-0.020	-0.047+
	[0.021]	[0.028]	[0.023]	[0.028]	[0.021]	[0.028]
Constant	0.426**	0.147*	0.147*	0.083	0.426**	0.147*
	[0.074]	[0.059]	[0.070]	[0.060]	[0.074]	[0.059]
Observations	1,205	1,205	1,205	1,205	1,205	1,205
R-squared	0.411	0.124	0.247	0.088	0.411	0.124
df	16	16	13	13	17	17

Note: Robust standard errors in brackets. +=significant at 10%; *=significant at 5%; **=significant at 1%

Figure 1: Common case of augmentation: AI provides recommendation for human working on the same task instance

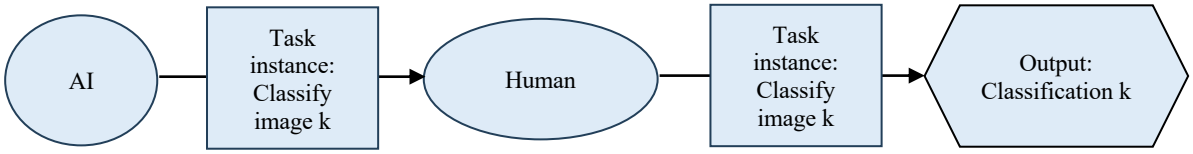


Figure 2: Common case of automation: Humans generate training data, AI then automates remaining task instances.

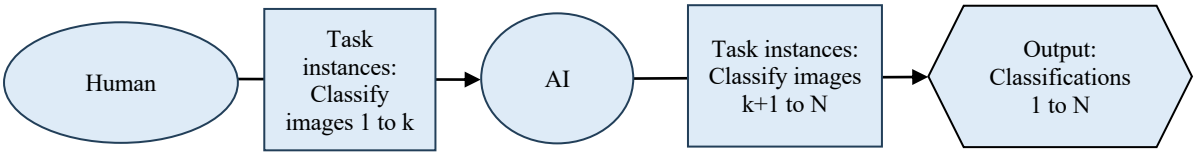
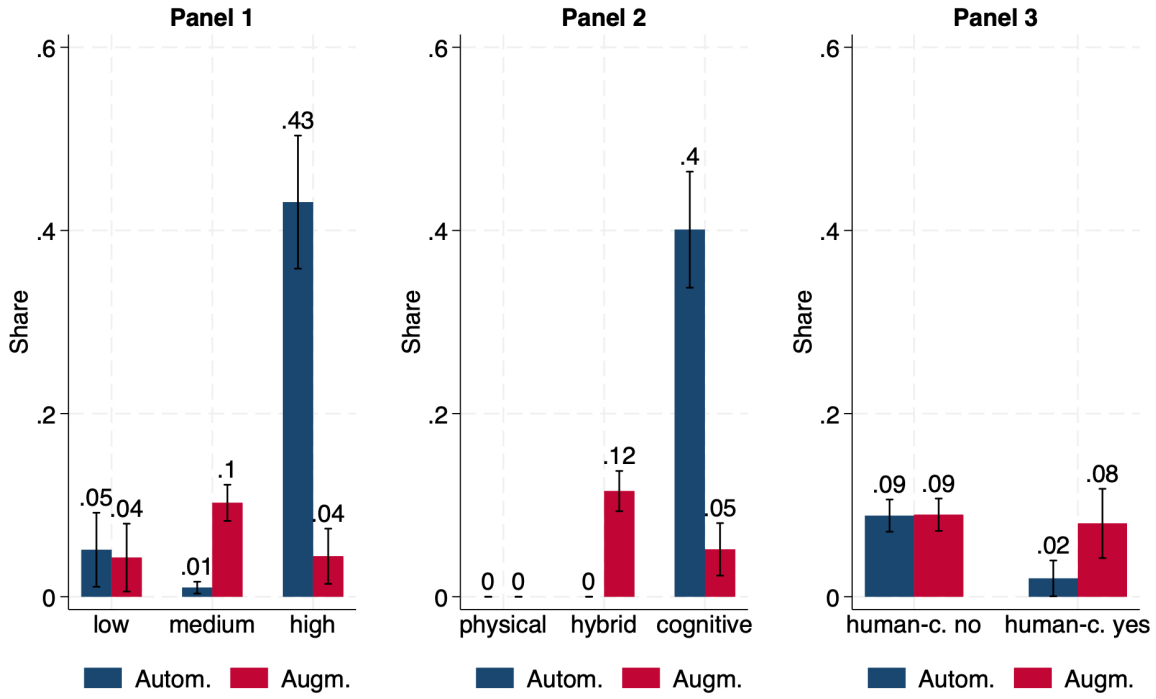


Figure 3: Share of projects using augmentation and automation, by task routinization (Panel 1), physical vs. cognitive nature of the task (Panel 2), and human-centered goals (Panel 3)



Note: Numbers above the 5% confidence intervals indicate the height of the bar.

Appendix A: Projects mentioned in text with link to 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
Butler County Stream Team	https://scistarter.org/butler-county-stream-team
DeVote	https://www.votemeanings.eu/
Earthwise Aware	https://www.earthwiseaware.org/
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
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/
Old Weather	https://www.oldweather.org/
Personal Genome Project	https://www.personalgenomes.org/

Phylo	https://phylo.cs.mcgill.ca/
Planet Hunters TESS	https://www.zooniverse.org/projects/nora-dot-eisner/planet-hunters-tess
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