Information Research - Vol. 30 No. iConf (2025)

Deskilling and upskilling with AI systems

Kevin Crowston and Francesco Bolici

DOI: https://doi.org/10.47989/ir30iConf47143

Abstract

Introduction. Deskilling is a long-standing prediction of the use of information technology, raised anew by the increased capabilities of AI (AI) systems. A review of studies of AI applications suggests that deskilling (or levelling of ability) is a common outcome, but systems can also require new skills, i.e., upskilling.

Method. To identify which settings are more likely to yield deskilling vs. upskilling, we propose a model of a human interacting with an AI system for a task. The model highlights the possibility for a worker to develop and exhibit (or not) skills in prompting for, and evaluation and editing of system output, thus yielding upskilling or deskilling.

Findings. We illustrate these model-predicted effects on work with examples of current studies of AI-based systems.

Conclusions. We discuss organizational implications of systems that deskill or upskill workers and suggest future research directions.

Introduction

The increased capability of modern artificial intelligence (AI) systems, generative AI in particular, has increased concerns about their impact. We define AI as '*systems that build on machine learning, computation, and statistical techniques, as well as rely on large data sets to generate responses, classifications, or dynamic predictions*' (Faraj, Pachidi, & Sayegh, 2018, p. 62). In this paper we focus on a long-standing concern about the impact of automation, namely deskilling, meaning that the work left for the humans requires a lower level of skill than the original job. AI raises the question of deskilling anew since as a general-purpose technology that could impact more kinds of work (Sison, Daza, Gozalo-Brizuela, & Garrido-Merchán, 2023). Consistent with the fear of deskilling, many AI applications are described as having a levelling effect, meaning that they help novices more than experts (i.e., levelling ability), which we interpret as deskilling. For instance, Brynjolfsson, Li, and Raymond (2023) found that a chatbot to support customer service workers enabled less experienced operators work at the level of more experienced one.

However, as Crowston and Bolici (2019) emphasize, automation does not always entail replacing human effort entirely. Instead, it often involves diverse patterns such as decision support or blended decision-making, where human expertise remains integral. This perspective highlights a more nuanced view of AI's role, suggesting that its impact on skills depends on how systems are designed and integrated into tasks. At the same time, some applications demonstrate that AI's benefits are not uniform across all users. While certain tools primarily support novices, others prove more powerful for experienced users, not causing deskilling but rather enhancing their skills. Indeed, some applications might even need new skills to use effectively, another form of upskilling.

The question we seek to address in this paper is, under what conditions do these two outcomes emerge? What are the characteristics of tasks that when automated in particular ways lead to a levelling effect of technology versus those where technology better supports more experienced users? This question is important to identify the implications for workers as AI capabilities are built into more systems. The answer also has implications for how organizations might staff functions using the system and the longer-term implications of system usage.

Literature review

A common and long-standing predicted effect of computerization is deskilling, meaning the replacement of skilled workers by those with less skill or reduced opportunities for the same workers to exercise particular skills. Concern about deskilling has been raised since the dawn of computing (Mann & Williams, 1960; Whisler, 1970). Use of computer systems can strip a job of its content, leaving only a dull routine. For example, instead of solving a problem, a worker might instead feed relevant data to a computer and have it solve the problem. As a result, workers lose the opportunity or time to develop their skills through experimentation or on-the-job learning, or even to maintain skills previously acquired (Ardichvili, 2022; Li, Zhang, Niu, Chen, & Zhou, 2023). For instance, Rinta-Kahila et al. (2023) found that a company's reliance on an accounting package with sophisticated automation rendered its accountants—and consequently the organization as a whole—unable to perform a specific accounting process without the software, which they refer to as skill erosion. Organizational disruption ensued when the software was replaced with another, less automated system.

Deskilling has knock-on effects for the nature of the work, which can reinforce skill loss. As the flow of work becomes more like an assembly line, an individual worker's pace becomes regulated by the needs of processes on either side, and the need for interaction and resulting opportunity for social ties are reduced. Glenn and Feldberg (1977) described this process as the '*proletarianization of clerical work*'. They noted that even fifty years ago, clerical jobs were becoming more like factory jobs, with increased subdivision of work and specialization of workers due to automation and use of scientific management principles from classic organization theory as management attempted to control workers and reduce the variability of their output. Zuboff (1988) pointed out that a system embodies assumptions about how the work should be done, resulting in a loss of flexibility for the worker. Formal rules replace discretion or specific knowledge, reducing workers' opportunities to display their mastery of their jobs. More recently, Holm and Lorenz (2022) found that when computers were used to give orders, the results for workers were increased work pace, constraints and decreased autonomy, an effect that was more pronounced for medium-skilled jobs. These changes in job content can lead to a loss of overview of the whole process (Ardichvili, 2022), further reducing workers' ability to learn and maintain appropriate skills.

The opposite prediction is upskilling. Computers can be used to automate the repetitive parts of a worker's job, leaving more interesting components for the human, and producing a more desirable job requiring a higher level of skills or having more responsibilities. For example, Zuboff (1988) presented a case in which the automation of a paper mill increased the role of the first-line production workers since they could control more than the single functions they used to. The jobs, therefore, required more skill, and the operators began to perform some of the functions of the managers. Even 60 years ago, Mann and Williams (1960) found some cases of job enlargement, noting that systems eliminated many routine jobs. Moreover, Sofia et al. (2023) (among many others) suggested that implementing AI will require new skills. They proposed that companies should help workers to identify which skills transfer and to develop needed new skills.

In practice, both effects, deskilling and upskilling, seem likely to occur simultaneously. There is some recent evidence from firm-level data of both effects. For example, Xue et al. (2022) found that Chinese companies reporting AI applications hire more employees without formal college education. However, McGuinness, Pouliakas, and Redmond (2023) found that skill-displacing technologies were positively associated with task variety and job-skill complexity, suggesting upskilling, though mostly for higher-skilled jobs. Zhang, Lai, and Gong (2024) also suggested an increase in employment for those with higher cognitive skills.

In past studies, deskilling or upskilling has often been viewed as dependent on deliberate choices about how to implement systems, driven by managers' preferences, e.g., for controlling versus working with workers (Zetka Jr, 1991). However, there is also a technical component to the decision as it is possible to design systems to promote hybrid intelligence (Rafner et al., 2022; Wahlström et al., 2024) and thus avoid deskilling. For example, Schemmer, Kühl, and Satzger (2022) described a decision support system that provides advice but requires users to make the final decision, thereby maintaining skill levels. Similarly, Arnold et al. (2023) designed a system with an interface based on expert knowledge representations and explanations, which improved novices' skills. And finally, the nature of the task itself is an important factor, interacting with managerial and technical impetuses.

Deskilling and upskilling due to AI

More recently, there have been a few studies that specifically include the differential impact of AI systems based on experience. We report on several that serve as the basis for our thinking. Table 1 summarizes these examples.

Brynjolfsson et al. (2023) reported on a study of an AI-based conversational assistant that supported the work of customer service agents by monitoring their chats with customers and suggesting possibly relevant documents to address the customers' problems. In a study with 5,179 customer support agents, they found that access to the tool boosted productivity, as indicated by a 14% increase in issues resolved per hour, while also increasing customer and worker satisfaction. However, the productivity increase was restricted to novice and low-skilled workers, who saw a 34% improvement; experienced and highly skilled workers experienced minimal benefit. They

Paper	Input	Evaluation	Editing	Impact
Brynjolfsson et al. (2023)	Extracted from customer chat	Relevance of document to problem	None	Deskilling
Dell'Acqua et al. (2023)	Prompt for problem to solve	Evaluation of suitability of suggestions	Output text lightly edited	Deskilling
Noy and Zhang (2023)	Used task prompts unchanged	Evaluate suitability of output	Output text lightly or not edited	Deskilling
Campero et al. (2022)	Prompt for UI element	Visual evaluation of appearance	Graphical interface to position; option to edit code	Deskilling
Peng et al. (2023) and Cui et al. (2024)	Prompt for needed code	Evaluate suitability of code	Edit code to fix bugs and adapt to need	Deskilling [a]
Luo et al. (2021)	Extracted from calls	Evaluate coaching advice	Implement coaching advice	Deskilling [b]
Wang et al. (2023)	Extracted from medical record	Determine suitability of billing code	Accept or reject code and possibly add others	Skill maintaining
Choudhury, Starr, and Agarwal (2020)	Extracted from patent application	Determine suitability of search terms	Possibly add additional search terms	Skill maintaining
Dell'Acqua (2024)	Extracted from the job application	Evaluate system suggestion	Accept or reject suggestion	Skill maintaining [c]
Kim and Kang (2024)	Preset inputs	Evaluate recommendation and important factors	Incorporate factors in report	Skill maintaining
Jia, Luo, Fang, and Liao (2024)	Preset inputs	None	None	Automation & upskilling

[a] Not statistically significant
[b] When volume of suggestions matched to user abilities
[c] With less accurate system

Table 1. Summary of results from the literature.

suggested that the AI model spreads the best practices of more proficient workers, which we interpret as evidence for deskilling because a worker need not be as skilled to perform well.

Dell'Acqua et al. (2023) reported on two experiments with Boston Consulting Group consultants. We focus on the first, in which 385 consultants carried out a set of 18 realistic consulting tasks designed to be within the capabilities of AI, namely, to conceptualize and develop new product ideas. Consultants either had no AI support, access to ChatGPT-4 or access to ChatGPT-4 with a prompt engineering overview. Consultants with access to ChatGPT were more productive and produced higher quality output, with a much stronger effect for consultants who performed lower on an initial assessment task. Those who received training in prompt engineering performed somewhat better than those who did not. These results show levelling and deskilling, as the system enables those who previously displayed less skill to operate at higher level.

In an experimental study with a writing task carried out by 453 college-educated professionals, Noy and Zhang (2023) found that those using ChatGPT both saved time and increased quality. Subjects with access to ChatGPT were given examples of prompts as a form of training. The authors reported that subjects seem to have mostly used ChatGPT's output as is, with little or no editing. Those who scored worse on an initial task improved their quality more, again evidence for levelling and deskilling.

Campero et al. (2022) explored having 200 programmers develop HTML code to replicate a web page, half using ChatGPT-3 with prior conditioning to generate relevant HTML code. The users had a graphical interface with which they could reposition the element and could edit the generated HTML if desired. The programmers using ChatGPT completed the task about 30% faster. Interestingly, when they had 50 non-programmers do the same task with ChatGPT, they found that 95% of them finished the task in about the same time as the professional programmers. They concluded that this use of AI can '*be seen as a form of deskilling for the programmers whose jobs could now be performed by people with less skill—and for lower compensation*'.

Peng et al. (2023) reported on a study of the productivity impacts of GitHub Copilot in which 95 programmers were recruited to write a simple HTTP server in JavaScript, 45 using Copilot and 50 without it. The treatment group finished in less than half the time, with roughly the same level of success. Though the effect was only marginally significant, they also found that developers with fewer years of experience benefited more, evidence again for deskilling.

Cui et al. (2024) reported on three experiments conducted by three large companies who randomly assigned developers to use CoPilot. They concluded that use of the tool leads to a 26% increase in weekly coding tasks completed over all three companies. Focusing just on Microsoft, they found that CoPilot adoption was higher for junior developers and developers with lower tenure in the company. Further, the increase in pull requests, commits and builds was roughly three times higher for lower tenure developers, though the differences were not statistically significant. Overall, these results seem consistent with deskilling.

Luo et al. (2021) reported on an AI coaching system for sales representatives. The system analyses the agents' calls to give advice about improving the interaction with customers. From an experiment with 429 agents, it was found that the system helped middle-ranked agents increase their sales rate the most, to nearly the level of higher-ranked agents. Lower-ranked agents were unable to absorb the volume of suggestions, while higher-ranked agents were averse to AI-generated advice. When the volume of suggestions was reduced, lower-ranked agents also improved, i.e., further levelling.

Wang, Gao, and Agarwal (2023) reported on the effects of an AI system to support coding of medical records. From a study with 80 coders using the system and 468 in the control group, they found that the system increased the productivity of all workers (reportedly with no impact on quality), but more so for those with more task experience, who could more quickly evaluate the proposed codes for suitability. They noted that '*If an AI is successfully trained on a task-specific data set, AI can substitute for a worker's task experience*'. However, in their study, the benefit went to those with more task experience. On the other hand, the system does not seem to have changed the nature of the skills required. In summary, this study seems to find neither deskilling nor upskilling, rather maintaining the advantage of having more experience.

Choudhury et al. (2020) examined how expert knowledge can address system shortcomings. Specifically, they studied a system that processes patent applications to suggest search terms to find relevant existing patents. Since patent applications may be deliberately designed to look different from the prior art, an experienced human patent examiner can complement the system by expanding the search. In an experiment, 221 MBA students examined five patent claims that were invalidate by an existing patent that used different language than the application. To simulate expertise, half of the subjects were given expert advice about how to search that included advice on adding terms. The experiment showed that those using the new system found a more precise set of relevant patents that were more like the application, as intended, but that the search was unlikely to find the relevant patent. The expert advice made it more likely that the patent would be found, again suggesting the importance of the existing expertise for this task.

Kim and Kang (2024) studied 97 mutual fund analysts who write reports rating mutual funds including an explanation for the rating. Half had access to a proprietary rating algorithm that rated the fund and identified the factors that were important for the prediction. They found that access to the predictions improved recommendation quality (i.e., whether the prediction matched the outcome) for simple cases, but had a negative impact on explanation quality, especially for junior analysts. New analysts were less likely to have algorithmic aversion but found it hard to incorporate the system results into their thinking and so wrote shorter reports that were less coherent and included more causal drivers in the explanation. We interpret this case as the system better supporting more skilled users, since skill was needed to make proper use of the system outputs.

In an experiment with recruiters evaluating job applications, Dell'Acqua (2024) varied the quality of the AI support provided. They found that expert recruiters using a less accurate system were more likely to carefully evaluate the applications themselves rather than simply taking the system suggestion, resulting in a more accurate evaluation than those using the more accurate (but imperfect) system without a careful evaluation. Indeed, more experienced recruiters using the better system performed worse than less experienced recruiters. They conclude that '*an AI that is 'too good' may induce workers to mindlessly follow algorithmic advice and lead to over-delegation*' and suggest that '*collaboration should be designed with the goal of keeping humans attentive in tasks where their focus is necessary to improve performance*'.

Finally, Jia et al. (2024) studied 40 sales agents interacting with 3144 potential customers to sell credit cards, working in two phases: first qualifying leads by assessing interest and then engaging to make a sale. Half of the agents used an AI telephone conversational system that autonomously did the first step, while the other half did it themselves. They found that agents using the AI system were more likely to make a sale because the system screened out likely-uninterested leads, allowing them to focus on better prospects. However, top agents were 2.8 times more likely to make a sale than bottom agents, which they attributed to the top agents' ability to develop better sales scripts and to answer questions for which they had not been trained, which bottom agents did not do. This case is evidence for upskilling: by taking over the routine part of a job, the system leaves work that requires more skill to perform at a high-level.

Model development

As a basis for analysing different applications of AI, we propose a simple model of the interaction among human, technology and task. In our model, the user performs a task that involves problem assessment and the creation of some output. For the scope of this paper, we focus on information tasks, not physical tasks, covering a broad category such as decision-making, customer care, brainstorming of ideas, etc. The model structure aligns with Crowston and Bolici's (2020) framework, which identifies three patterns of machine learning use—decision support, blended decision-making, and complete automation—and highlights how automation can affect not only specific tasks but also interdependent processes and coordination mechanisms. When using a system to support a task, rather than performing the task directly, users follow a process including: 1) assessing the task that should be executed, 2) possibly formulating an input and providing it to the AI system, 3) assessing the result, 4) accepting, regenerating, or editing the output, and 5) completing the task. For example, a human interacting with a document repository to find an answer to a problem will formulate a query (or use a query generated by the system), look at results to see if they meet the requirements, pick one or redo the query and try again. For interaction with a large language model (LLM) such as ChatGPT, the human will formulate a prompt, evaluate the generated results, tweak the prompt if the results are unsatisfactory, and possibly edit the output to improve it to complete the task.

Model components

Understanding the deskilling or upskilling impacts of AI requires a comprehensive model that captures the interaction between four main elements: 1) Humans, 2) Systems, 3) the Outputs generated, and 4) the Tasks that must be performed in the organization by humans and/or system. It is based on the previous research on the roles of expertise, prompting, system accuracy and task nature in assessing AI-enabled work. For example, Zuboff (1988) observed that systems reflect work assumptions, limiting flexibility and eroding skills (Human–Task; Human–AI), while Rinta-Kahila et al. (2023) demonstrated skill erosion from over-reliance on automation (Human–AI; Human–Task). Conversely, application of hybrid intelligence (Schemmer, Kühl and Satzger, 2022) mitigates deskilling by involving humans in decision-making (Human–AI; Human–Task). Brynjolfsson et al. (2023) illustrated a leveling effect, enabling novices to reach intermediate performance, though not expert levels (Human–Outputs; Human–Task). Wang, Gao, and Agarwal (2023) emphasized human expertise in refining AI outputs, underscoring the dual-edged nature of AI's impact on skills and performance (Human–Outputs; Human–AI). Our model is intended to analyse these interactions and their implications for skill development, use and retention. The proposed model highlights the interplay between human expertise, system capabilities, and task requirements in shaping task performance outcomes.

1) *HUMAN*: the persons that must perform the task and who are deciding if and how to use a system to support or to substitute for their work. We focus on two main characteristics that have an impact on the process:
    a) Domain knowledge: the extent to which the human is an expert in the specific domain relevant to the task. Higher domain knowledge enables better assessment and refinement of system outputs.
    b) Input formulation knowledge: the human's ability to effectively formulate inputs for the AI system. For instance, when using an LLM, expertise in prompting can significantly influence the quality and relevance of the system's outputs.

2) *AI*: the specific system that can be accessed by the human during the task execution and for which we consider three characteristics:
    a) Input variability: The variability of input to the system. Some systems take a fixed set of variables while in contrast, an LLM can take nearly any text as input.
    b) Accuracy/limitations: The constraints of the system, such as the propensity for generating errors or the need for human intervention to correct and refine outputs.

3) *OUTPUTS* of the system: the answers that the system provides in response to the input that have at least the following characteristics:
    a) Quality: The accuracy, relevance, and usability of the system-generated outputs. High-quality outputs require less modification and are more useful for completing the task.
    b) Speed: We presume that the system will be able to generate an answer more quickly than the human, leading to the observed increases in speed.

4) *TASK*: the activity that must be performed and for which we can consider:
    a) Nature of the task: the specific characteristics of the task, including whether it is creative, analytical, or procedural.
    b) Task division: how the task is split between the human and the system. This could involve the human performing the entire task, the system performing the entire task, or a collaborative effort where both the human and system contribute.

Model phases

Considering the model as a representation of the dynamics of interaction between the various elements of the process, we can distinguish four different temporal phases.

1) *Phase One*: the human responsible for carrying out the task can decide whether to be supported by a system and, if so, in what way. It may be though that usage is non-discretionary, meaning the human user is obligated to use the system.

2) *Phase Two*: the human utilizes their domain knowledge and skills to interact with the system to obtain support. It may be that the inputs to the system are predetermined by the task, or the user may have freedom to craft an input.

3) *Phase Three*: the results generated by the system are assessed and interpreted by the human, who decides whether to accept them or refine either manually or through further through additional interaction with the system.

4) *Phase Four*: the final results are used to execute the task, either in support of or as a substitute for direct human involvement.

Interaction among the components

The interplay between the components of the model determines the overall impact on skill levels and task performance.

- *Human and System Interaction*: the effectiveness of the system may be dependent on the human's ability to craft an input. Those with high knowledge can generate better initial responses from the system or be better in refining the outputs iteratively tuning the prompting itself. Indeed, it may be that the prompts are created by experts who develop a system rather than by the end-user using the system for a task.

- *System and Output*: the system's capabilities and limitations directly affect the quality and adaptability of the outputs. High-quality outputs reduce the need for extensive human intervention and can be applied to a variety of tasks, enhancing productivity.

- *Human and Output*: the human's role in assessing and interpreting the system's output depends (again) on their domain knowledge. High domain knowledge allows for quicker and more accurate assessment of the output, reducing the risk of simply accepting a wrong or incomplete result. Experts in the domain can better refine system outputs if they are not suitable.

- *Outputs and Task*: the nature of the outputs influences how the task is performed. High-quality, adaptable outputs can enhance productivity and potentially upskill workers by allowing them to focus on higher-level refinements. On the other hand, poor outputs can lead to deskilling if the human's role is reduced to merely accepting or rejecting system-generated content without substantial engagement.

This phased approach highlights the iterative and interactive nature of the model, emphasizing the crucial role of human expertise at each stage to maximize the effectiveness of the LLM and ensure the successful completion of the task. The conclusion of the model can lead to different impacts on the need for expertise.

- *No Effect*: in this scenario, the use of AI has no impact on the skills of the individuals involved. The task is performed similarly whether or not the AI is used, and the human's existing knowledge and skills remain unchanged. However, the system may have other benefits, e.g., for speed or quality.

- *Levelling Effect*: this scenario occurs when AI minimizes the importance of the human's knowledge on task performance. The use of AI flattens the importance of prior knowledge, as a novice using AI can achieve a task performance like that of an expert. In this case, the AI levels the playing field, reducing the skill gap between novices and experts.

- *Multiplier Effect*: in this scenario, the use of AI acts as a multiplier on the human's existing knowledge, thereby increasing the performance gap between novices and experts. The AI enhances the capabilities of those with higher prior knowledge, leading to significantly better task performance compared to novices. This effect underscores the role of AI in amplifying the skills and expertise of experienced users.

By understanding these different scenarios, we can better anticipate the implications of AI integration into various workflows and design strategies to optimize both human and AI contributions to task performance.

We speculate that a system with pre-formed prompt with results that are easy to assess and that have little need to edit more likely results in levelling and so deskilling, as a more expert worker does not have an opportunity to employ their expertise. On the other hand, a system could have more flexibility about prompting or more need for output assessment and editing, tasks that experts could be quicker and more accurate in doing. To the extent that the task has these characteristics, it is more likely to benefit from expertise and thus help experts more than non-experts.

Findings: deskilling and upskilling due to AI

We illustrate our model by analysing some of the studies surveyed above. For instance, in Brynjolfsson et al. (2023)'s study, the prompt is taken from the customer chat, not the agent. The agent needs to assess if a proposed document is apropos but can also suggest it and let the customer assess. If appropriate, the solution is provided to customer as is. Therefore, our model suggests that the effect of the system will be levelling, as found: the system can provide solutions that a more experienced employee would suggest, but without requiring the same level of expertise. Similarly, in Noy and Zhang (2023)'s study, subjects using ChatGPT seem to have copied the writing prompts from the problem and used ChatGPT's output largely unchanged. They had to evaluate if the output was suitable but given the similarity of the task to their regular work, we expect this evaluation to be straightforward (that is, subjects differed in the quality of their writing, but we think not in the ability to assess suitability of output).

In about half of our examples, the input to the system was generated from the task, eliminating the need or ability of a user to develop expertise in directing the system. However, these cases include our two examples of skill maintenance. In other cases, users reportedly used the given task prompts unchanged, again not developing skills. In two of the cases we reviewed, users could prompt more freely (Dell'Acqua et al., 2023; Peng et al., 2023). Interestingly though, those cases still resulted in deskilling. In summary, the cases reviewed do not suggest that the ability to better prompt an AI system is as yet a distinguishing characteristic in task performance.

In contrast, ability to evaluate and make use of a system's output does seem to play a role in several cases. Wang et al. (2023)'s study poses an interesting example. In this case, the search is based on sentences in the medical record. However, the authors report that evaluation of the suggestions was required to rule out false positives, which was quicker for more experienced workers. In this case, the system does not require new skills (e.g., for prompting) and maintains the value of existing skills (evaluation). In Kim and Kang (2024)'s study, expertise was seemingly needed to evaluate the system's output and to incorporate it into the final product, making the system more useful for more experienced workers. Choudhury et al. (2020) found that the search terms identified by the system needed to be augmented, which required expertise to do successfully. Luo et al. (2021) found an inverted U-shape effect for the initial coaching system: it did not help experienced workers, who already knew the job, nor inexperienced workers who could not cope with the volume of suggestions, but did help people in between, the later effect again highlighting the importance of being able to assess and incorporate system output.

Finally, the case of Jia et al. (2024) is one of few studies we found that reportedly resulted in upskilling. Interestingly, this case is also one in which one subtask was completely automated, namely the initial screening call with a potential customer, while the remaining subtask left to the human is performed without support, a subtask for which greater skill translates into better performance. (In other words, the analysis in Table 1 does not describe the task the human performs.) We are curious what the impact would be of supporting the sales task and speculate that it could lead to levelling, as found by Luo et al. (2021).

Overall, we perceive a general pattern: if you have too little skill, you can't make use of the system outputs. If you have moderate skill, the system generally seems to help achieve better performance. If you have a lot of skill, the system doesn't help as much and may even be resisted (Wang et al., 2023).

Discussion

The model and the studies reviewed more broadly suggest several points for consideration.

First, reviewing the papers identified, we note a lacuna, namely we identified no studies in which better prompting skills gave more experienced workers an advantage. We expected that using Copilot to support programming would have these effects. For instance, Mozannar et al. (2024) observed that programmers using CoPilot spent over 20% of their time thinking about or verifying a CoPilot suggestion, about 10% of the time editing a suggestion, and about 10% crafting prompts. Prompt crafting is often iterative: write a prompt, assess output, tweak the prompt. Often, suggestions were accepted to fully evaluate and tweak them, not necessarily because they were correct. Dibia et al. (2022) found that experienced programmers still found incorrect code suggestions from CoPilot useful even if the code was not entirely correct, as it could be modified with little effort, thereby increasing productivity. Similarly, Zamfirescu-Pereira et al. (2023) found that while the code generated by CoPilot often had errors, they were easier to fix than errors in code generated by humans. They concluded that '*CoPilot can become a liability if it is used by novice developers who may fail to filter its buggy or non-optimal solutions due to a lack of expertise*' (Zamfirescu-Pereira et al., 2023). Randazzo et al. (2024) suggests that ChatGPT users who retain overall control of the task, strategically deciding which tasks to delegate, perform better than those who direct the system through the whole task and much better than those who delegate entirely. However, we only found two studies (Cui et al., 2024; Peng et al., 2023) that examined the impacts of individual differences using this technology and unfortunately, these studies do not provide much detail about how developers interacted with the system.

Second, questions about deskilling and upskilling have important organizational implications that need to be considered. For instance, if the system results in deskilling, organizations may be tempted to hire less skilled workers or to invest less in training since performance with the system will still be satisfactory. These temptations will likely be greater for jobs that face high turnover, such as customer support. A consideration is that managers tend to systematically underestimate the expertise needed to do the work of their employees, meaning that they may classify more work as replaceable or low skilled than is appropriate. This consideration reinforces the importance of involving the people doing the work in system design. A further consideration is the implications for organizational learning. If the problem is not static, but the system has a levelling effect, then who will learn the answers to the new questions, if there are no longer any experts doing the tasks? Relatedly, it is an open question whether non-experts using a system learn to do the task independently or whether the system obviates the need to learn.

On the flip side, systems that reward expertise also raise concerns. If expertise is more valued, organizations need to consider how it is developed. For instance, there are anecdotal reports of companies no longer hiring entry-level workers to do what LLMs can do (Edwards, 2023; Yegge, 2024). If the work of entry-level positions can be largely automated, organizations will face the problem of how new hires develop the necessary expertise to oversee the automated work.

Third, system use may require user to develop new skills in prompt crafting and in evaluation and use of system output, rather than manual creation of output. There is some evidence for these effects, e.g., the small improvements found by Dell'Acqua et al. (2023) for the short prompt-crafting training, and the several studies in which expertise was needed to evaluate outputs (Choudhury et al., 2020; Kim & Kang, 2024; Luo et al., 2021; Wang et al., 2023). However, Dell'Acqua (2024) raise the issue of needing to motivate workers to be critical about systems, more critical ironically for systems that work better.

Fourth, the future of work with AI and the related necessary skills requires consideration of the inherent nature of AI, which is best able to provide answers for problems and solutions that frequently appear in its training data. This limitation could lead to a need for workers who have skills primarily in identifying corner cases and their possible solutions that the AI cannot handle. Understanding and designing how to support these types of skills in workers remains an open question. If the management (identification of problems and solutions) of the most common cases is done through or with AI, it cannot represent a learning field for new workers who will need to learn to handle specific and less frequent corner cases.

Fifth, our model also posits limits to the impacts of technology support. Specifically, Amdahl's law (Amdahl, 1967) says that speed up due to a new system component is limited by fraction of time the using new component. As an example, if only 10% of a job is automated, the maximum speed-up is 1/90% or about an 11% speed-up. Reasoning in reverse, to make someone 2x faster at their work (i.e., a multiplier effect), as found by Peng et al. (2023) for the programmers using ChatGPT, requires eliminating 50% of what they do. We speculate that such a result implies that programmers are seeing benefits by having the system write entire functions at a time, rather than writing lines of code. Our model does not as yet capture the possibly transformative effects of entirely changing the nature of the task performed.

Finally, our model has at least one design implication. As prompting is a new skill that has a potential to make a difference to the results, it might be beneficial to let users tweak the prompts, even if they are mostly preset or derived from the task data. The visibility may help people to develop expertise in prompting and use this new skill to improve results.

Conclusion

We conclude with some ideas for future research. First, this model is based on examination of a few sample implementations of AI to support work. More systematic studies across a wider range of tasks would help refine it and demonstrate its utility. To carry out these studies will require more details about the nature of the task, the technology and the workers' interactions. It would be helpful to have more detail about the specifics of the systems and how people interact. We would like to dig into the details of the system more to understand where skills make a difference. For instance, it could be that crafting a good query for a search is a more important skill than getting an LLM prompt exactly right, given the latter's flexibility and interpretive abilities.

Second, use of the model could guide studies of new systems. For instance, it would be interesting to vary the level of prompt crafting possible for the same task and exploring impact on workers with varying skill levels. We expect more-skilled workers to use the capabilities to further extend their advantage over less-skilled workers, but this is an empirical question that needs study.

Third, it is important to consider that this model, like most studies focused on AI to date, focused on the relationship between work and AI within a single task. However, work is a process made up of multiple interdependent tasks. Therefore, the potential impact of AI on work should be examined within the context of how AI is used across a set of tasks that need to be coordinated. This broader perspective acknowledges that the integration of AI affects not only individual tasks but also the overall workflow, requiring a broader approach to understand its full implications on job performance and skill requirements.

Acknowledgements

Kevin Crowston was partly supported by a grant from the United States National Science Foundation, Grant 21-29047.

About the authors

Kevin Crowston is a Professor of Information Science in the School of Information Studies, Syracuse University, Syracuse, New York, USA. He received his Ph.D. from the Massachusetts Institute of Technology and his research interests are in the futures of work using advanced technologies. He can be contacted at crowston@syr.edu

Francesco Bolici is a Professor of Organization Studies and Information Systems and Director of OrgLab at the Department of Economics and Law, Università degli studi di Cassino, Italy. Holding a Ph.D. from Luiss Guido Carli University, Rome, his research examines how information technology enables new possibilities for reimagining work design and organizational practices. He can be contacted at f.bolici@unicas.it

References

Amdahl, G. M. (1967, April 18-20). Validity of the single processor approach to achieving large scale computing capabilities. In Proceedings of the Spring Joint Computer Conference (pp. 483–485).

Ardichvili, A. (2022). The impact of artificial intelligence on expertise development: Implications for HRD. Advances in Developing Human Resources, 24 (2), 78–98. https://dx.doi.org/10.1177/15234223221077304

Arnold, V., Collier, P. A., Leech, S. A., Rose, J. M., & Sutton, S. G. (2023). Can knowledge-based systems be designed to counteract deskilling effects? International Journal of Accounting Information Systems, 50, 100638. https://dx.doi.org/10.1016/j.accinf.2023.100638

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative ai at work (Tech. Rep. No. Working Paper 31161). National Bureau of Economic Research. Campero, A., Vaccaro, M., Song, J., Wen, H., Almaatouq, A., & Malone, T. W. (2022). A test for evaluating performance in human-computer systems (https://arxiv.org/abs/2206.12390).

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. https://dx.doi.org/10.1002/smj.3152

Crowston, K. & Bolici, F. (2019) Impacts of machine learning on work. In Proceedings of the Hawaii International Conference on System Sciences, Hawai'i, USA. https://dx.doi.org/10.24251/HICSS.2019.719

Crowston, K., & Bolici, F. (2020). Impacts of the Use of Machine Learning on Work Design. In Proceedings of the 8th International Conference on Human-Agent Interaction (pp. 163-170). https://doi.org/10.1145/3406499.3415070

Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2024). The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. Social Science Research Network (SSRN). https://dx.doi.org/10.2139/ssrn.4945566

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (Tech. Rep. No. 24-013). Harvard Business School Technology & Operations Management Unit. (https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf)

Dell'Acqua, F. (2024). Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters. Laboratory for Innovation Science, Harvard Business School. (https://www.fabriziodellacqua.com/)

Dibia, V., Fourney, A., Bansal, G., Poursabzi-Sangdeh, F., Liu, H., & Amershi, S. (2022). Aligning offline metrics and human judgments of value of AI-pair programmers (https://arxiv.org/abs/2210.16494).

Edwards, B. (2023). IBM plans to replace 7,800 jobs with AI over time, pauses hiring certain positions. (https://arstechnica.com/information-technology/2023/05/ibm-pauses-hiring-around-7800-ro)

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28 (1), 62-70. https://dx.doi.org/10.1016/j.infoandorg.2018.02.005

Glenn, E. N., & Feldberg, R. L. (1977). Degraded and deskilled: The proletarianization of clerical work. Social Problems, 25 (1), 52–64.

Holm, J. R., & Lorenz, E. (2022). The impact of artificial intelligence on skills at work in Denmark. New Technology, Work and Employment, 37 (1), 79–101.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67 (1), 5–32.

Kim, H., & Kang, X. (2024, Aug). Machine predictions and causal explanations: Evidence from a field experiment. Symposium presentation at the Academy of Management Meeting, Chicago, IL, USA. https://dx.doi.org/10.5465/AMPROC.2024.13661symposium

Li, C., Zhang, Y., Niu, X., Chen, F., & Zhou, H. (2023). Does artificial intelligence promote or inhibit on-the-job learning? Human reactions to AI at work. Systems, 11 (3), 114.

Luo, X., Qin, M. S., Fang, Z., & Qu, Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. Journal of Marketing, 85 (2), 14-32. https://dx.doi.org/10.1177/0022242920956676

Mann, F. C., & Williams, L. K. (1960). Observations on the dynamics of a change to electronic data-processing equipment. Administrative Science Quarterly, 217–256.

McGuinness, S., Pouliakas, K., & Redmond, P. (2023). Skills-displacing technological change and its impact on jobs: Challenging technological alarmism? Economics of Innovation and New Technology, 32 (3), 370–392.

Mozannar, H., Bansal, G., Fourney, A., & Horvitz, E. (2024). Reading between the lines: Modelling user behaviour and costs in AI-assisted programming. In Proceedings of the CHI Conference on Human Factors in Computing Systems. https://dx.doi.org/10.1145/3613904.3641936

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381 (6654), 187–192.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot (https://arxiv.org/abs/2302.06590).

Rafner, J., Dellermann, D., Hjorth, A., Veraszto, D., Kampf, C., MacKay, W., & Sherson, J. (2022). Deskilling, upskilling, and reskilling: A case for hybrid intelligence. Morals & Machines, 1 (2), 24–39.

Randazzo, S., Lifshitz-Assaf, H., Kellogg, K., Dell'Acqua, F., Mollick, E. R., & Lakhani, K. R. (2024). Cyborgs, centaurs and self automators: Human-genAI fused, directed and abdicated knowledge co-creation processes and their implications for skilling. (https://ssrn.com/abstract=4921696)

Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W., & Ruissalo, J. (2023). The vicious circles of skill erosion: A case study of cognitive automation. Journal of the Association for Information Systems, 24 (5), 1378–1412.

Schemmer, M., Kühl, N., & Satzger, G. (2022). Intelligent decision assistance versus automated decision-making: Enhancing knowledge workers through explainable artificial intelligence. In Proceedings of the 55th Hawai'i International Conference on System Sciences. Virtual. (https://hdl.handle.net/10125/79517)

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). ChatGPT: More than a "weapon of mass deception": Ethical challenges and responses from the human-centered artificial intelligence (HCAI) perspective. International Journal of Human–Computer Interaction, 1–20.

Sofia, M., Fraboni, F., De Angelis, M., Puzzo, G., Giusino, D., Pietrantoni, L. et al. (2023). The impact of artificial intelligence on workers' skills: Upskilling and reskilling in organisations. Informing Science: The International Journal of an Emerging Trans-discipline, 26, 39–68.

Wahlström, M., Tammentie, B., Salonen, T.-T., & Karvonen, A. (2024). AI and the transformation of industrial work: Hybrid intelligence vs double-black box effect. Applied Ergonomics, 118, 104271.

Wang, W., Gao, G., & Agarwal, R. (In press). Friend or foe? Teaming between artificial intelligence and workers with variation in experience. Management Science. https://dx.doi.org/10.1287/mnsc.2021.00588

Whisler, T. L. (1970). Information Technology and Organizational Change. Belmont, CA: Wadsworth.

Xue, M., Cao, X., Feng, X., Gu, B., & Zhang, Y. (2022). Is college education less necessary with AI? Evidence from firm-level labor structure changes. Journal of Management Information Systems, 39 (3), 865–905.

Yegge, S. (2024). The death of the junior developer. (https://sourcegraph.com/blog/the-death-of-the-junior-developer)

Zamfirescu-Pereira, J., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://dx.doi.org/10.1145/3544548.3581388

Zetka Jr, J. R. (1991). Automated technologies, institutional environments, and skilled labor processes: Toward an institutional theory of automation outcomes. Sociological Quarterly, 32 (4), 557–574.

Zhang, W., Lai, K.-H., & Gong, Q. (2024). The future of the labor force: Higher cognition and more skills. Humanities and Social Sciences Communications, 11 (1), 1–9.

Zuboff, S. (1988). In the Age of the Smart Machine: The Future of Work and Power. Basic Books, Inc.