Deskilling and upskilling with generative AI systems

Kevin Crowston
Syracuse University
School of Information Studies
crowston@syr.edu

Francesco Bolici
Università degli studi di
Cassino e del Lazio Meridionale
f.bolici@unicas.it

Abstract

Deskilling is a long-standing prediction of the use of information technology, raised anew by the increased capabilities of generative AI (GAI) systems. A review of studies of GAI applications suggests that deskilling (or levelling of ability) is a common outcome. We propose a model of a human interacting with a GAI system for a task that suggests settings more likely to yield deskilling vs. upskilling. 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. We illustrate these effects with examples of current studies of GAI-based systems. We discuss organizational implications of systems that deskill or upskill workers and suggest future research.

1. Introduction

The increased capability of modern artificial intelligence (AI) systems, generative AI (GAI) in particular, has increased concerns about their impact. There is lot of fear about automation replacing workers, even to the point of leaving people with no work and in need of sources of income other than employment. However, it seems more likely that AI will more often be used to support workers, doing some tasks while humans do others (Bolici and Crowston, 2019). The question then is the impact of having some of the tasks in a job supported or even performed by an automated system while the human worker performs the others.

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. GAI raises the question of deskilling anew since as a general-purpose technology that could have broader impact on more kinds of work (Sison et al., 2023). Consistent with the fear of deskilling, many GAI 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 et al. (2023) found that a chatbot to support customer service workers enabled less experienced operators work at the level of more experienced one. But other applications seem to be more powerful for more experienced users, i.e., not deskilling but rather skill enhancing. Indeed, some advanced 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 GAI capabilities are built in to more systems. The answer also has implications for how organizations might staff functions using the system and the longer-term implications of system usage.

2. 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 has been raised since the dawn of computing (e.g., Mann and Williams, 1960; Whisler, 1970). Computer systems can strip a job of its content, leaving only a dull routine. For example, instead of directly 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 et al., 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. An organizational disruption ensued when the software was replaced with another, less automated one.

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 clerk'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) describe this process as the "proletarianization of clerical work". They note 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 attempts to control workers and reduce the variability of their output. Zuboff (1988) points 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) presents 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) suggest that implementing AI will require new skills. They propose 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. Attewell and Rule (1984) report both deskilling and upgrading, and noted that it is difficult to determine which predominates. 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 et al. (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 et al. (2024) also suggest 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 (Wahlström et al., 2024; Rafner et al., 2022) and thus avoid deskilling. For example, Schemmer et al. (2022) describe 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.

In summary, while automation can lead to deskilling, resulting in job simplification and loss of expertise, it can also lead to upskilling, enhancing job complexity and requiring higher levels of cognitive skills. The net effect depends on how organizations choose to implement these technologies and the strategies they adopt to balance automation with human expertise. We seek to explore how the particular capabilities of GAI and how it is implemented will affect the degree of deskilling.

2.1. Deskilling and upskilling due to AI

More recently, there have been a few studies specifically about the differential impact of generative AI systems based on experience. We report on several that serve as the basis for our thinking.

Brynjolfsson et al. (2023) report 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 suggest that the AI model spreads the best practices of more proficient workers, which we interpret as evidence for deskilling because a worker does not need to be as skilled to perform at a high level.

Dell'Acqua et al. (2023) report 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 reportedly 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. Again, these results provide some evidence of levelling and deskilling, as the system enables those who previously displayed less skill to operate at higher level.

Peng et al. (2023) report 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, which might be viewed as evidence for deskilling.

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 report 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 conclude that this use of GAI 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".

Wang et al. (2023) reported on the effects of an AI system to support coding of medical records (though this is not GAI strictly speaking). 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. On the other hand, workers with longer tenure in the organization (uncorrelated with task experience) were more likely to find problems with the system and so resist its use. They note 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 actually 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.

Jia et al. (2024) did a study with 40 sales agents interacting with 3144 potential customers to sell credit cards, 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 attribute 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.

3. Model development

As a basis for analyzing different applications of generative 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. When using a system to support a task, rather than performing the task directly, users follow

Table 1. Summary of results from literature

Paper	Prompt	Evaluation	Editing	Outcome
Brynjolfsson et al. (2023)	Derived from customer chat	Relevance of document to problem	None	Deskilling
Dell'Acqua et al. (2023)	Crafted by user. Some benefit from a prompt crafting class	Necessary	Output text lightly edited	Deskilling
Peng et al. (2023)	Crafted by user	Necessary	Necessary	Deskilling?
Noy and Zhang (2023)	Reportedly used task prompts unchanged	Necessary	Output text lightly or not edited	Deskilling
Campero et al. (2022)	Conditioning to create HTML code; human prompt for HTML element	Visual evaluation of appearance of element	Graphical interface to position; possible to edit code but not necessary	Deskilling
Wang et al. (2023)	Extracted from medical record	Required	Accept or reject code and possibly add others	Skill maintaining
Jia et al. (2024)	Set by programmers	None	None	Upskilling

a process including: i. assessing the task that should be executed, ii. formulating a prompt, iii. assessing the result, iv. accepting, regenerating or editing the output, and v. completing the task. For example, a human interacting with a document repository to find an answer to a problem will formulate a query, look at results to see if they meet the requirements, pick one or redo the query and try again. For interaction with a GAI system, specifically 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.

3.1. Model components

Understanding the deskilling or upskilling impacts of GAI requires a comprehensive model that captures the interaction between for main elements: 1. Humans, 2. LLMs, 3. the Outputs generated, and 4. the Tasks that have to be performed in the organization by humans and/or GAI. Each element has specific characteristics and interacts with the others in complex ways. Our model is designed to analyze these interactions and their implications for skill development, use and retention.

1. **HUMAN**: the persons that have to perform the task and that are deciding if and how to use a GAI to support or to substitute their work. We focus on two main characteristics that have an impact on the. process:
   - *Domain Knowledge*: This refers to 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 LLM outputs.
   - *Prompting Knowledge*: This measures the human's ability to effectively formulate prompts for the LLM. Expertise in prompting can significantly influence the quality and relevance of the LLM's outputs.

2. **LLM (Large Language Model)**: the specific system that can be accessed by the human during the task execution and for which we consider two main characteristics:
   - *Proficiency/Capability*: The inherent abilities of the LLM to generate content based on the provided prompts. This includes understanding context, generating coherent responses, and the ability to refine answers iteratively.
   - *Accuracy/Limitations*: The constraints of the LLM, 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 GAI provides as results of the prompting, that have at least the following characteristics:
   - *Quality*: The accuracy, relevance, and usability of the LLM-generated outputs. High-quality outputs require less modification and are more useful for completing the task.

- *Speed*: We presume that the LLM will be able to generate an answer more quickly than the human, leading to the observed inceases in speed.

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

3.2. 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 an LLM and, if so, in what way.

2. **Phase Two**: The human utilizes their domain knowledge and prompting skills to interact with the LLM to obtain support (Figure 1).

3. **Phase Three**: The results generated by the LLM are assessed and interpreted by the human, who decides whether to accept them or refine further through additional prompting (Figure 2).

4. **Phase Four**: The final results are used to execute the task, either in support of or as a substitute for direct human involvement (Figure 3).

3.3. Interaction among the components

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

- *Human and LLM Interaction*: The effectiveness of the LLM is highly dependent on the human's prompting knowledge. Those with high prompting knowledge can generate better initial responses from the LLM or be better in refining the outputs iteratively tuning the prompting itself.

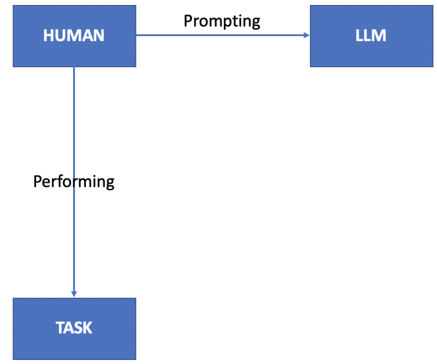

Figure 1. Phase 2: human accesses a LLM to get support for a task

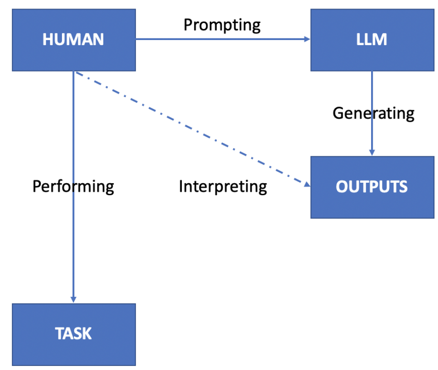

Figure 2. Phase 3: LLM produces results to the prompt that are interpreted by human

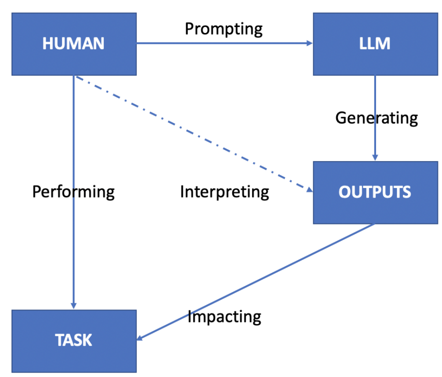

Figure 3. Phase 4: results interpreted by human are used in assessing the task

Indeed, it may be that the prompts are created by experts who develop a system rather than by the end-user using the LLM for a task.

- *LLM and Output*: The LLM'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 LLM'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 LLM 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 LLM-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.

1. **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.
2. **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 GAI can achieve a task performance similar to that of an expert. In this case, the AI levels the playing field, reducing the skill gap between novices and experts.
3. **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 GAI 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. 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."

4. Findings: Deskilling and upskilling due to AI

We illustrate our model by analyzing some of the studies surveyed in section 2.1. 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). Campero et al. (2022)'s study also somewhat fits our model. Subjects could prompt freely but the system was conditioned to generate HTML in response to simple cues (e.g., "add a green button"). Evaluation could be done simply by examining the result and edits could be made graphically as well as by editing HTML, so not requiring particular expertise.

Wang et al. (2023)'s study poses an interesting contrast to the previous studies. In this case (which is not strictly speaking GAI), 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).

Finally, the case of Jia et al. (2024) is the one study 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 providing some support for the sales task and speculate that it could lead to levelling. Relatedly, Luo et al. (2021) report on an AI coaching system for sales representatives. The system analyzes 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. However, lower-ranked agents were unable to absorb the volume of suggestions, while higher-ranked agents were adverse to AI-generated advice. When the volume of suggestions was reduced, lower-ranked agents also improved, i.e., further levelling.

Reviewing the papers identified, we note a lacuna, namely we identified no studies in which better prompting and evaluation skills gave more experienced workers an advantage. We expected that using Copilot to support programming would have these effects, but did not find any studies other than Peng et al. (2023) that examined the impacts of individual differences.

5. Discussion

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

First, it is 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 later's flexibility and interpretive abilities.

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 deskilled 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?

On the flip side, systems that reward expertise also raise concerns. First, system use may require those with expertise to develop new skills in prompt crafting rather than manual creation of output. There is some evidence for this benefit, e.g., the small improvements found by Dell'Acqua et al. (2023) to the short prompt crafting training. Second, if expertise is more valued, organizations need to consider how it is developed. If the work of entry-level positions can be largely automated, organizations may face the problem of how new hires develop the necessary expertise.

Additionally, the future of work with AI and the related necessary skills requires consideration of the inherent nature of AI, which tends 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. 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.

Our model also posits limits to the impacts of technology support. Specifically, Amdahl's law (1967) says that speed up due to a new system component is limited by fraction of time 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, 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 can 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.

6. Conclusion

We conclude with some ideas for future research. First, this model is based on examination of a few sample implementations of GAI 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.

Second, use of the 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 composed 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.

References

Amdahl, G. M. (1967). Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the Spring Joint Computer Conference*, pages 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.

Arnold, V., Collier, P. A., Leech, S. A., Rose, J. M., and Sutton, S. G. (2023). Can knowledge based systems be designed to counteract deskilling effects? *International Journal of Accounting Information Systems*, 50:100638.

Attewell, P. and Rule, J. (1984). Computing and organizations: What we know and what we don't know. *Communications of the ACM*, 27(12):1184–1192.

Bolici, F. and Crowston, K. (2019). Impacts of machine learning on work. In *Proceedings of the Hawai'i International Conference on System Science*, Wailea, HI.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative ai at work. Technical Report Working Paper 31161, National Bureau of Economic Research.

Campero, A., Vaccaro, M., Song, J., Wen, H., Almaatouq, A., and Malone, T. W. (2022). A test for evaluating performance in human-computer systems. Technical Report arXiv:2206.12390.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Technical Report 24-013, Harvard Business School Technology & Operations Management Unit.

Dibia, V., Fourney, A., Bansal, G., Poursabzi-Sangdeh, F., Liu, H., and Amershi, S. (2022). Aligning offline metrics and human judgments of value of AI-pair programmers. Technical Report arXiv:2210.16494.

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

Holm, J. R. and 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., and Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, 67(1):5–32.

Li, C., Zhang, Y., Niu, X., Chen, F., and 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., and Qu, Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing*, 85(2):14–32.

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

McGuinness, S., Pouliakas, K., and 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., and Horvitz, E. (2024). Reading between the lines: Modeling user behavior and costs in AI-assisted programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI ’24.

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

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. Technical Report arXiv:2302.06590.

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

Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W., and 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., and Satzger, G. (2022). Intelligent decision assistance versus automated decision-making: Enhancing knowledge workers through explainable artificial intelligence. In *Proceedings of the 55th Hawaii International Conference on System Sciences*. Virtual.

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., and 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*, pages 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 Transdiscipline*, 26:39–68.

Wahlström, M., Tammentie, B., Salonen, T.-T., and 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., and Agarwal, R. (2023). Friend or foe? Teaming between artificial intelligence and workers with variation in experience. *Management Science*.

Whisler, T. L. (1970). *Information technology and organizational change*. Wadsworth, Belmont, CA.

Xue, M., Cao, X., Feng, X., Gu, B., and 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.

Zamfirescu-Pereira, J., Wong, R. Y., Hartmann, B., and 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*, pages 1–21.

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., and 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.