Information and Organization 35 (2025) 100559

Contents lists available at ScienceDirect

Information and Organization

journal homepage: www.elsevier.com/locate/infoandorg

Novice risk work: How juniors coaching seniors on emerging technologies such as generative AI can lead to learning failures

Katherine C. Kellogg [a,*], Hila Lifshitz [b], Steven Randazzo [b], Ethan Mollick [c], Fabrizio Dell'Acqua [d], Edward McFowland III [d], François Candelon [e], Karim R. Lakhani [d]

[a] MIT Sloan School of Management, United States of America
[b] Warwick Business School, Artificial Intelligence Innovation Network, United Kingdom
[c] The Wharton School, University of Pennsylvania, United States of America
[d] Digital Data Design Institute, Harvard Business School, United States of America
[e] Seven2, BCG Henderson Institute, France

A R T I C L E I N F O

Keywords:
Artificial intelligence
Experts
Generative AI
Risk work
Novices
Professional organizations
Professionals
HCI
Communities of practice
Work and occupations
Technology and organizations

A B S T R A C T

Historically, junior professionals have mentored senior professionals around new technologies, because juniors are typically more willing than seniors to perform lower-level tasks to learn new skills, better able than seniors to engage in real-time experimentation close to the work itself, and more willing than seniors to learn innovative methods that conflict with traditional identities and norms. However, we know little about what happens when emerging technologies have a high level of uncertainty in their use, because they have wide-ranging capabilities and are exponentially changing. With the rise of Artificial Intelligence, specifically learning algorithms and LLMs, such contexts may be increasingly common. In our study conducted with the Boston Consulting Group, a global management consulting firm, we interviewed 78 junior consultants in July–August 2023 who had recently participated in a field experiment that gave them access for the first time to generative AI (GPT-4) for a strategic business problem solving task. Drawing from junior professionals' in situ reflections soon after the experiment, we found that junior professionals may fail to manage risks around uncertain emerging technologies because juniors are likely to recommend three kinds of *novice risk work* tactics that: 1) are grounded in a lack of deep understanding of technologies that have uncertain and wide-ranging capabilities and are changing exponentially, 2) focus on change to human routines rather than system design, and 3) focus on interventions at the project-level rather than system deployer- or ecosystem-level. The implications of *novice risk work* are that, when junior professionals are expected to be a source of expertise in the use of uncertain, emerging technologies, this can lead to learning failures. This study contributes to our understanding of occupational learning around emerging technologies, risk work in organizations, and human-computer interaction.

* Corresponding author.
E-mail address: kkellogg@mit.edu (K.C. Kellogg).

https://doi.org/10.1016/j.infoandorg.2025.100559
Received 30 May 2024; Received in revised form 1 February 2025; Accepted 5 February 2025
Available online 21 February 2025
1471-7727/

1. Introduction

In the contemporary environment of accelerating technological change, senior professionals must quickly implement emerging technologies today (e.g., Cameron & Rahman, 2022; Karunakaran, 2021; Lebovitz et al., 2022; Mazmanian et al., 2013; Rahman & Valentine, 2021; Waardenburg et al., 2022). At the same time, they must anticipate future versions of technologies, and the implications of these technologies for both their clients and their own organizations (e.g., Bailey & Leonardi, 2015; Barley, 1986; Barrett et al., 2012; Endacott & Leonardi, 2022; Karunakaran et al., 2022; Lebovitz et al., 2021).

The literature on learning within occupational communities tells us that, when rapid technological change and shifts in the nature of work undermine existing expertise (e.g., Barley & Kunda, 2001; Beane, 2019; Bechky, 2020; Bharatan et al., 2022; Leonardi & Bailey, 2008; Sergeeva et al., 2020), this poses a problem for senior members of communities of practice who have traditionally engaged in situated learning over time to become highly skilled in the use of the technologies their work requires (e.g., Bailey & Barley, 2011; Brown & Duguid, 1991; Lave & Wenger, 1991; Orr, 1990). Under these conditions, senior members of an occupation may

Table 1
Novice Risk Work.

	Novice Risk Work	Expert Risk Work
Understanding of Capabilities	**Novices' Suggestions for Managing Risks Reflect Misunderstanding of the Technology's Capabilities**	**Experts' Suggestions for Managing Risks Reflect Understanding of the Technology's Capabilities**
Accuracy	● Use a standardized way of asking questions ● Do the work first without GenAI	● Decide on appropriate use cases where error risks are acceptable ● Test GenAI's reliability in executing each subtask
Explainability	● Explain model logic to managers ● Agree on practices for explainable output	● Avoid GenAI use where a high degree of explainability is required ● GenAI provides illusion of transparency, but explanations may not match true actions
Contextualization	● Use only for cases where contextualization is not necessary	● Provide contextual information, and specify the desired output ● Use RAG to add content
Target of Change	**Novices Suggest Managing Risks by Changing Human Routines**	**Experts Suggest Managing Risks by also Changing Data, Model, and System Design**
Accuracy	● Train users to validate results ● Ask managers to review all user prompts/responses	● Set up second automatic monitoring system to check if in line with users' goals ● Use a model that provides links to sources ● Use a more accurate model
Automation Complacency	● Train users to take ownership of work when using GenAI	● Design a system that provides proactive self-reflective prompts ● Build an interface that visualizes uncertainty ● Build a default prompt that prompts the user to input their goals; Support pattern-matching between GenAI suggestions and users' task goals ● Apply hard-coded restrictions
Contextualization	● Train users in prompt engineering (also an expert tactic) ● Agree on cases where contextualization is not necessary	● Improve prompts centrally and build them into the system ● Design system to begin with a user prompt to communicate goals
Level of Intervention	**Novices Suggest Managing Risks by Intervening at the Project Level**	**Experts Suggest Managing Risks by also Intervening at the System Deployer- and Ecosystem-Level**
Accuracy	● Have managers and users on project agree on conditions under which GenAI can be used reliably ● Have managers review users' work process	**System Deployer Level** ● Provide co-audit tools ● Communicate to users the intended conditions under which GenAI can be used reliably ● Create a prompt library of effective prompts for particular tasks ● Continually assess the alignment of LLMs vis-à-vis evaluation metrics ● Establish feedback/incident reporting mechanisms **Ecosystem Level** ● Assess credibility of data sources; use trusted sources; Perform real-time data updates ● Flag/correct misleading responses ● Explain system's capabilities and limitations ● Avoid anthromorphizing the AI
Automation Complacency	● Ask managers to give users adequate time ● Ask managers not to short sell cases	**System Deployer Level** ● Effectively onboard users; Provide personalized adjustments for users ● Apply cognitive forcing functions ● Allow users to clearly specify human-GenAI task allocation **Ecosystem Level** ● Attend to the impact on users of a model's style, tone, or perceived personality ● Highlight prompt changes and the resulting output changes ● Support debugging by generating test cases and test data to help identify corner cases

attempt to develop practical expertise in the use of new technologies by learning from junior professionals (e.g., Barley, 1986; Beane & Anthony, 2024; Kaplan et al., 2017; Kellogg et al., 2021; Leonardi, 2007). Senior professionals often attempt to learn from junior professionals, because junior professionals are typically more willing to perform lower level tasks to learn new skills (Beane, 2024; Kellogg, 2009; Pratt et al., 2006). Junior professionals are also better able to engage in real-time experimentation with new technologies, because juniors don't risk losing their mandate to lead if others realize that they have little practical expertise (e.g., Beane, 2019; Beane & Anthony, 2024; Endacott & Leonardi, 2022; Leonardi, 2007). Finally, junior professionals are more willing to learn new methods that conflict with existing identities (e.g., Lifshitz, 2018; Nelson & Irwin, 2014), practices (e.g., Beane & Orlikowski, 2015; Mazmanian & Beckman, 2018; Pachidi et al., 2021) and frames (e.g., Anthony & Tripsas, 2016; Mazmanian, 2013).

However, this literature has not explored "emerging technologies" such as artificial intelligence, data analytics, social media, digital platforms, blockchain, and 3-D printing (e.g., Arora et al., 2023; Polykarpou et al., 2020; Rahman et al., 2024)—technologies treated as emerging "because their uses and effects are still varied and have yet to stabilize around a recognizable set of patterns and because the technologies themselves are, by design, always changing and adapting" (e.g., Bailey et al., 2022, p. 1). Such emerging technologies present novel risks (e.g., Barrett et al., 2024; Hardy & Maguire, 2020) because of their uncertain and wide-ranging capabilities, exponential rate of change, potential for outperforming humans in a wide variety of skilled and cognitive tasks, and dependence on a vast, varied, and high volume of data and other inputs from a broad ecosystem of actors (e.g., Schneider et al., 2024). Contexts in which juniors professionals may gain access to the technologies that have uncertain and wide-ranging capabilities that are exponentially changing may be increasingly common. Thus, we investigate the research question: How do junior professionals deal with emerging technologies that present novel risks, and what implications does this have for occupation members learning to use these technologies?

We have unique data on junior professionals adopting GenAI (an emerging technology) in its nascent stage (first few months of Gen AI at work) in a professional organization (The Boston Consulting Group). These unique data enable us to detail the ways that junior professionals suggest managing novel risks associated with emerging technologies, and the potential impacts on occupation members learning to use the technologies.

In the sections that follow, we elaborate the concept of *novice risk work*, which highlights that junior professionals may fail to manage risks around nascent emerging technologies because juniors 1) lack a deep understanding of technologies that have uncertain and wide-ranging capabilities and are changing exponentially, 2) have daily work experience with human-to-human interaction that leads them to focus on change to human routines rather than system design, and 3) have daily work experience at the project level that leads them to focus on interventions at the project level rather than at the system deployer- or ecosystem-level (Table 1). Because of junior professionals' *novice risk work*, juniors coaching seniors in effective technology use may lead to learning failures within occupational communities around uncertain emerging technologies.

2. Expertise, learning, and intra-occupational dynamics with the introduction of new technologies

Research on learning to use technologies in occupational work most often examines how junior professionals develop expertise around an occupation's core technologies by learning from senior experts (e.g., Bailey & Barley, 2011; Beane, 2024; Nicolini et al., 2022; Orr, 1990). Junior professionals typically learn from senior experts to use an occupation's technologies by engaging in legitimate peripheral participation as the juniors progress up the ranks of seniority (e.g., Brown & Duguid, 2001; Hutchins, 1995; Lave & Wenger, 1991). When new technologies are introduced, this poses a considerable challenge to traditional intra-occupational dynamics.

New technologies can dramatically change how tasks are performed (e.g., Bechky, 2020; Beunza & Stark, 2004; Chan & Hedden, 2023; Leonardi & Bailey, 2008; Nelson et al., 2023) and change the kinds of problems that experts can solve (e.g., Huising, 2015; Treem & Leonardi, 2017; Zetka, 2003). To adapt to new technologies, senior professionals must learn new methods (Anthony, 2018; Leonardi, 2007; Nelson & Irwin, 2014). Yet, doing so is extremely difficult for senior professionals (Bailey & Leonardi, 2015), because such new methods require changing existing practices and frames (Anthony & Tripsas, 2016; Fayard et al., 2016; Lee et al., 2020; Pachidi et al., 2021), and enacting new professional roles across boundaries (Barley, 1990; Beane & Orlikowski, 2015; Bechky, 2003; Pine & Mazmanian, 2015; Sergeeva et al., 2020).

Under such conditions, senior professionals may attempt to learn from junior professionals, rather than the other way around (e.g., Beane & Anthony, 2024). Juniors may be better able than their senior counterparts to learn to effectively use the new technologies for several reasons. First, junior professionals are often closest to the work itself, because they are the ones engaging in concrete and less complex tasks (e.g., Barley, 1996; Bharatan et al., 2024; Karunakaran, 2018; Kellogg, 2009; Pine & Mazmanian, 2017; Rahman & Barley, 2017). Second, junior professionals are usually more able to engage in real-time experimentation with new technologies, because they do not risk losing their mandate to lead if those around them, including clients as well as those more junior to them, recognize that they lack the practical expertise to support their hierarchical position (e.g., Beane, 2019; Beane & Anthony, 2024; Endacott & Leonardi, 2022; Leonardi, 2007). Third, junior professionals may be more willing to learn new methods that conflict with existing identities (e.g., Lifshitz, 2018; Nelson & Irwin, 2014), practices (e.g., Beane & Orlikowski, 2015; Mazmanian & Beckman, 2018) and frames (e.g., Anthony & Tripsas, 2016; Mazmanian, 2013).

For example, when CT scanners arrived in radiology, senior radiologists learned to effectively use them from junior technicians who were closest to the work of injecting dyes and constructing images (Barley, 1986). When help-desk queuing technology arrived in IT, senior technicians responding to help-desk calls began to ask for help from junior technicians who had engaged in experimentation with the new technology and developed skill in its use and features (Leonardi, 2007). And when patient referral and tracking technology arrived in primary care, senior medical assistants learned from those more junior how to follow each colonoscopy patient's progress through procedure and follow-up, because the juniors were more willing to learn new methods that conflicted with existing identities, norms, and frames (Kellogg et al., 2021).

In our study of BCG consultants, as the existing literature would anticipate, the junior professionals we interviewed expected that senior professionals would turn to juniors to educate seniors in the effective use of the new technology. Yet, contrary to what the current literature would expect, the *novice risk work* tactics that the juniors recommended to manage the risks that GenAI posed to seniors' valued outcomes ran counter to the risk management tactics recommended by experts in GenAI technology at the time. This revealed that the junior professionals coaching seniors would lead to learning failures around this uncertain emerging technology.

3. Bringing novel risks, risk work, and novice tactics into our understanding of learning around uncertain emerging technologies

We found organization theory's concepts of *novel risks* and *risk work*, and human-computer interaction (HCI) theory's concept of *novice tactics for technology use* to be helpful to understanding the dynamics we observed.

3.1. GenAI presents novel risks

Scholars of risk in organizations (e.g., Barrett et al., 2024; Corvellec, 2010), suggest that organizations cannot identify and manage risk if the risk is novel—"the probability, nature, and magnitude of its adverse effects cannot be ascertained through prevailing scientific knowledge and practices" (Hardy & Maguire, 2020, p. 685). GenAI presents novel risks that are different from risks posed by the technologies that have been examined by scholars in the existing literature on junior professionals coaching seniors (Table 2).

First, *GenAI can be accessed and customized by novice users without coding and without owning infrastructure* (e.g., Schneider et al., 2024). Prior studies of juniors teaching seniors have analyzed technologies in which interacting with AI required infrastructure and raised barriers for ordinary users. For example, the CT scanning technology that Barley (1986) studied required a large investment in hardware by hospital administrators, and integration with existing hospital medical equipment and IT infrastructure by IT professionals. In contrast, GenAI allows novice users to access the technology directly from their computers and collaborate with it in a nearly instantaneous fashion.

Second, *GenAI has uncertain and wide-ranging capabilities and is changing at an exponential rate* (e.g., Achiam et al., 2023; Webb et al., 2023; Wei et al., 2022). Prior studies of juniors teaching seniors have examined technologies purpose-built for specific applications, and changing at a slower rate. For example, the Davinci robot that Beane (2019) studied was purpose-built to allow a single surgeon to perform both support (retraction) and direct surgical action (dissection) to an unprecedented degree. In contrast, with GenAI, a broad range of applications can be performed by a single system, and GenAI's capabilities are expanding exponentially.

Third, *GenAI carries the possibility of outperforming humans in a wide variety of skilled and cognitive tasks* (e.g., Bubeck et al., 2023). Prior studies of juniors teaching seniors have investigated technologies designed to assist professionals. For example, the EMR

Table 2
Potential Novel Risks Presented by Uncertain Emerging Technologies.

	Technologies Studied in Existing Literature	Uncertain Emerging Technologies
Similar Sources of Risk		
Historical relationship between positional and practical expertise	Position and title corresponded with knowledge and skill, which is cumulative across roles	Similar
Disruption of new technology to existing knowledge and skill by distancing senior members from their work	Adopting new technologies may mean replacing embodied methods with teleoperated methods Analytical technologies may bring a a shift from direct, intentional analysis to distanced, encoded analysis	Similar
Produces results that are partly beyond the control of the user organization	Technology may gather data directly from external sources, automatically calculate key metrics, and feed these calculations directly into organizational tools	Similar
Output is often difficult to understand	It is often difficult for users to comprehend how the technology works	Similar
Data may lead to biases in decisions of AI systems	Technology may be trained on imbalanced data that lead models to exhibit, for example, a higher error rate for some demographic groups than others	Similar
Potential Novel Sources of Risk		
Simplicity and infrastructure required for access and usage by novice users	Interacting with the technology requires infrastructure and raises barriers for ordinary users	The technology can be accessed and customized by novice users without coding and without owning infrastructure
Range of applications	A single system is often purpose built to execute a particular task	A single system can perform a broad range of applications
Rate of change	The technology is quickly evolving through product life cycle enhancements and user modifications	The technology is changing at an exponential rate
Has possibility of outperforming humans in a wide variety of skilled and cognitive tasks	Technology helps to inform and supplement human action	The technology has unprecedented, often superhuman performance to act more autonomously and conduct more tasks
Source of data for the technology	Data used in the system is often generated within a company	Data used in the system stems increasingly from a mix of various entities that dynamically interact

technology that Kellogg et al. (2021) studied was able to assist medical assistants to do their work of patient check-in processes and referrals and insurance authorizations more accurately and quickly. In contrast, GenAI holds the potential to surpass human performance across a diverse range of skilled and cognitive activities.

Fourth, *GenAI combines data and other inputs at an unprecedented scale and detail from a broad ecosystem of actors* (e.g., Arora et al., 2023). Prior studies of juniors teaching seniors have analyzed technologies that involved model developers, system deployers, and system users, but were dependent on a limited set of data from each. For example, the Factset and CapIQ technology that Anthony (2021) studied gathered data directly from public SEC filings, automatically calculated particular key metrics, and then fed these calculations directly into spreadsheets in the user's organization. In contrast, with GenAI, continuous change in data sources at an unprecedented scale and detail affect what the technology is able to do. Thus, GenAI emerges through a set of expanded relations and continues to emerge in new ways as those relations evolve.

3.2. *Risk work: The identification and management of risk objects*

In the context of novel risks such as these, organizational actors may engage in "risk work" (Power, 2016a)—designating an object (such as GenAI technology) as risky, and prescribing actions to manage the "risk object" (e.g., Barrett et al., 2024). Identification of a risk object reflects subjective preferences in the sense that a risk exists when and only when someone perceives the possibility of someone or something being harmed (e.g., Campbell, 2006). Therefore, risk is identified as actors discuss values at stake in a particular setting and suggest that a "risk object" (such as GenAI technology) poses new dangers or harm to something that is valued in that setting (such as accuracy or explainability of outcomes).

Risk work is thus a creative act that is constrained by the procedures, routines, mental schemes, and strategies of justification of an actor's everyday practice (Corvellec, 2010). For example, because junior professionals work in a hierarchical relationship with their managers, they are likely to identify values at stake that they know are important to their managers. In addition, since junior professionals' everyday practice involves human routines, and project-based work, their suggested measures to control novel risks will likely be shaped by these aspects of their everyday practice.

Risk work practiced by experts in an emerging technology will similarly reflect the values important in their setting, and these experts will similarly suggest risk management measures related to their everyday practice. For example, compared to junior professionals, experts in generative AI will have had a higher degree of exposure to GenAI systems. In addition, because of their daily practice, the experts will be more oriented to systems and models rather than human routines, and to system deployer and ecosystem effects rather than only project-level effects. In situations like the one we studied, in which junior professionals had just gained access to an emerging technology that had a high level of uncertainty in its use, because it had wide-ranging capabilities and was exponentially changing, experts may be better positioned to manage novel risks than are junior professionals.

Actors such as the junior professionals and GenAI experts we studied are likely to act on novel risks by recommending action on other risks with which they are more familiar (Hardy & Maguire, 2020). Risk work thus transforms a novel risk object (such as GenAI technology) into a known risk object (such as a project-based human routine) that poses some kind of familiar organizational risk that can be assessed and managed through tried and tested techniques, thereby rendering it actionable (Hardy & Thomas, 2014).

3.3. *Novice tactics for technology use*

Risk work conducted by junior professionals, in addition to being constrained by aspects of their everyday practice, may be also constrained by their position as novices in the use of GenAI technology. Theorists of human-computer interaction highlight that novice programmers often intuitively approach interactions with a new technology in counterproductive ways (e.g., Ko et al., 2004; Lahtinen et al., 2005). These theorists show, for example, that novice end users may bring human intuitions rooted in social experiences to technology use that lead these novices to attempt debugging opportunistically rather than systematically (e.g., Ko et al., 2004). Novices may also attempt to design an ML model by directly mapping a personal need to a model task rather than by framing an achievable task (e.g., Yang et al., 2018). Surfacing pitfalls that non-experts are susceptible to can help to inform the design of technologies that are easier to effectively use by people who are not formally trained in their use (e.g., Yang et al., 2018; Zamfirescu-Pereira et al., 2023). HCI literature on novice end users suggests that designers can build future tools that better support novice end users in the face of novices' common counterproductive intuitions and struggles.

Taken together, the concepts of *novel risks*, *risk work* and *novice tactics for technology use* help to explain why junior professionals learning to use an uncertain emerging technology may intuitively approach interactions with GenAI technology in counterproductive ways and may identify risks and prescribe actions to manage them by targeting human routines at the project level. Indeed, this is what we found in our study of BCG consultants. We further found that such risk work can lead to learning failures, because it does not allow junior professionals to identify and manage the novel risks posed by an uncertain emerging technology such as GenAI.

In what follows, we first briefly describe our methodological approach. We next elaborate how the junior professionals we studied expected that they would need to coach seniors in the use of GenAI technology, yet also identified GenAI technology as posing new dangers to valued outcomes in their setting. We then describe the actions that junior professionals prescribed to manage the novel risks associated with GenAI, and compare their recommended measures to those recommended by GenAI experts in the emerging literature at the time. This comparison highlights how junior professionals' risk work around emerging technologies may be constrained by juniors' position as novices in the use of technologies that have uncertain and wide-ranging capabilities that are exponentially changing and by juniors' everyday practices involving human routines at the project level; juniors' *novice risk work*, in turn, can lead to learning failures around uncertain emerging technologies.

4. Brief description of methodological approach

Here we provide a brief description of our methodological approach. A more detailed description of our methods is provided in **Appendix A.**

The findings presented here are derived from 78 60-min interviews of Boston Consulting Group consultants (junior professionals) conducted via Zoom after they had participated in a field experiment in which they used generative AI (GPT-4) to complete a problem-solving task. Consultants highlighted for us that they expected that they would need to coach their managers (senior professionals) in the use generative AI. They also expected that their managers would perceive particular risks associated with the technology, and that they, as junior professionals, would need to manage these risks in order to coach these senior professionals in the use of GenAI technology. Consultants discussed tactics that they thought could address these risks.

We noted that the tactics the consultants identified to manage risks associated with GenAI appeared to be different from tactics to manage the same set of risks that were being recommended by experts in generative AI writing at the time. Thus, we searched on google scholar to gather articles by generative AI experts (written up until the time of our interviews with the consultants) about how to best address the set of risks identified by the consultants—risks to accuracy, explainability, contextualization of outputs and of automation complacency of GenAI users. We compared consultants' suggested risk work tactics to address managers' concerns about these valued outcomes to those that were being recommended by experts in generative AI at the same time. This allowed us to highlight differences in risk work tactics prescribed by novices versus experts in an uncertain emerging technology.

5. New barrier to junior professionals coaching seniors: GenAI as a risk object

5.1. Junior professionals expected that they would need to coach seniors in the use of GenAI technology, but did not expect characteristics associated with seniors' position to be a barrier

As the existing literature on communities of practice would lead us to anticipate, the junior professionals we interviewed expected that they would likely be the ones to educate the senior professionals about how to effectively use the new technology. For example, juniors noted:

Jnr36: "We'll need to teach managers [to use GenAI]…to try to educate them."

Jnr31: "I think the biggest thing for managers is us just training them….We need to show managers that [GenAI] is not a magic wand waving solution. Maybe having them even just watch how an associate would do a single module, or have an understanding of what that looks like, would help them kind of right size in their head what [GenAI] can do."

Jnr66: "Whenever I discover something cool, I will share it with my managers. Like the new [GenAI] Powerpoint tool."

Surprisingly, given the current literature on juniors coaching seniors in the use of new technologies, we did not find that juniors expected that threats to seniors' position, seniors' need to focus on complex work, or seniors' commitment to existing norms and frames would be key obstacles to junior professionals successfully coaching more senior professionals in the effective use of new technology. Two juniors said:

Jnr38: "It's just about creating a dialogue with them about what you are doing. And it's about making them feel more comfortable with it, through trial and error and training."

Jnr36: "We just need to put the reasoning in front of their faces."

Additional examples of juniors' expectations around the need to coach seniors are shown in **Appendix B**.

5.2. Instead, juniors' designated GenAI as a risk object, and identified novel AI risks as the key barrier to juniors coaching seniors

Instead, juniors noted that the key obstacle to their ability to coach seniors in effective use of the new technology would be the *novel risks that the technology posed* to outcomes that seniors valued. Juniors highlighted that seniors would be concerned that GenAI posed a risk to the *accuracy of outputs* (the degree to which outputs have attributes that correctly reflect the true value of the intended attributes of a concept or event in a particular context of use; Steimers & Schneider, 2022). Two juniors related:

Jnr24: "Obviously the first [thing managers will worry about is that] leveraging GenAI can produce output that is actually incorrect. So that would be a big concern."

Jnr3: "Managers will be scared that [new consultants] would bring something made up…If a [consultant] is doing desk research on new technology around vitamins and minerals, he could ask GPT, and ask for the links. Since he doesn't know how consulting works, he might use the links directly from GPT, and give it to the manager."

Juniors also suggested that seniors would be concerned about risks to the *explainability of outputs* (the degree to which the outputs are presented in a way that is understandable for humans; Steimers & Schneider, 2022). Two juniors said:

Jnr7: "Managers will be worried about GPT [because] it's a black box."

Jnr6: "Using GenAI will [negatively] affect a lot of the communication between consultants and managers, because consultants used to be able to back up everything, and now they won't be able to."

Juniors further noted that seniors would be concerned about risks to the *contextualization of outputs* (the degree to which outputs are coherent, relevant, or in compliance with particular constraints or rules due to absorbing and taking into account pertinent contextual data when producing outputs; Sai et al., 2024). Two juniors said:

Jnr24: "[Managers will] be concerned about the tool being able to take into account all the variables or context that are actually necessary to put together a valid solution…It's all about context. If you are working in a very nuanced industry with a very specific set of guidelines for the client, having that knowledge is one thing, and being able to feed that into a tool to actually spit out valid or actionable recommendations is another."

Jnr63: "Managers may be concerned about, 'Is GPT output specific enough [given the context of the case and the client's problem]?'…Like, if it's for a client that has a lot of their financial information that isn't available online, [then managers would be concerned about] is [GPT output] being tailored and specific enough to the client's context as opposed to the generic publicly available information."

Finally, juniors expected that seniors would be concerned that GenAI posed a risk to user engagement by promoting *automation complacency* (in which users may trust the outputs provided by generative AI in situations where they should not; Van Dis et al., 2023). Two juniors said:

Jnr12: "Managers will worry about complacency. They'll want to make sure that consultants aren't just blindly accepting GPT output."

Jnr63: "I think there will be a trust issue, [with managers wondering], 'Did you do this yourself, or did you use a tool to do it? Did you really think through that it was the right approach? Have you double checked it? Did you reread it through, and edit it to better fit our context? Did you look at any other sources to check if this is accurate information?' All those questions of, did you do your due diligence on this work that someone else effectively did for you."

Additional examples of juniors' expectations around key barriers to coaching of seniors being the novel risks that GenAI posed are shown in **Appendix C**.

6. Novice risk work

We compared juniors' suggested AI risk management tactics to those recommended by GenAI experts in the emerging literature at the time. We found that the juniors we studied recommended tactics for managing GenAI risks that differed in three key ways from tactics recommended by GenAI experts at the time. Junior professionals' *novice risk work tactics* included: 1) *tactics that stemmed from a lack of deep understanding of the characteristics of the uncertain emerging technology*, 2) tactics informed by juniors' everyday practice including human-to-human interaction that *overestimated the potential of changing human routines (rather than system design)* to mitigate these risks, 3) tactics informed by juniors' everyday practice of project-level work that *overestimated the potential of making changes at the project level (rather than the deployer- or ecosystem-level)* to mitigate these risks (Table 1). We elaborate each of these tactics below, and compare them to tactics suggested by GenAI experts at the time. The tactics are summarized in Table 1, and additional examples of juniors' descriptions of each of the tactics are presented in **Appendix Table D1-D3**.

6.1. Novice risk work type #1: Lacking deep understanding of the characteristics of the uncertain emerging technology

The first type of novice tactics we discovered were juniors' recommendations for managing GenAI output risks related to accuracy, explainability, and contextualization, which demonstrated juniors' lack of a deep understanding of the capabilities of GenAI technology. Juniors' lack of deep understanding is not surprising since, at the time of the study, these juniors were novice users. They were not technical experts, and the technology had a high level of uncertainty in its use, because it had wide-ranging capabilities and was exponentially changing.

Yet, it is important to understand this type of risk work, because the current literature does not examine contexts in which professionals need to learn to use uncertain emerging technologies; and such contexts are increasingly common. Some examples of this type of novice risk work are given below, with additional examples shown in **Appendix D, Table D1**.

6.1.1. Some juniors lacked a deep understanding of GenAI accuracy

Some consultants did not understand that the accuracy of GenAI was ultimately limited at the time of the study. Thus, these juniors recommended managing the risk of GenAI output inaccuracy in two ways. First, they suggested using a standardized way of asking questions:

Jnr19: "We could have a standardized way for asking questions in GenAI, and a standardized way of summarizing the AI steps, and make that available to everyone."

Second, they suggested that juniors do the work themselves first, before using GenAI:

Jnr24: "GenAI tools should be used later on in the process. I think that when you're working on that initial problem-solving component, you should be doing it yourself, manually. And then once you have more of a finalized output, you can cross-check, or use generative AI to be able to add on any kind of additional thoughts you might have. So, we could leverage that approach. Do the work yourself first. Only use Gen AI once you have any kind of preliminary output, and augment only versus create."

In contrast, articles by GenAI experts available at the time demonstrated that generative AI models could confidently present users with information that was hallucinatory, and that this could result in incorrect output that did not accurately reflect real people, places, or facts (e.g., Weidinger et al., 2022). Hallucinations could occur when the model tried to fill in gaps in its knowledge or when the input was ambiguous. Thus, experts recommended addressing GenAI's accuracy issues by deciding on appropriate use cases where error risks were acceptable (e.g., Achiam et al., 2023), and by independently testing GenAI's reliability in executing each subtask; for example by breaking down the users' needed subtasks (e.g., "information gathering") and creating evaluations for each independently (Shavit et al., 2023).

6.1.2. Some juniors lacked a deep understanding of GenAI explainability

Some consultants did not understand that the explainability of GenAI was not possible at the time of the study. Thus, these juniors recommended two tactics for addressing this risk. First, they suggested explaining GenAI's rationale to managers:

Jnr52: "The manager may question what GenAI did, so you need to be able to explain it."

Jnr41: "It's about understanding the source of the recommendation or the result. Being able to explain it."

Second, they suggested agreeing on practices for explainable output:

Jnr19: "You could be careful in asking the AI how it got to the answer, not sure how well it works. We could have a standardized way of summarizing the AI steps."

In contrast, articles by GenAI experts available at the time demonstrated that LLMs did not provide transparency for their reasoning (e.g., Jain & Wallace, 2019; Jacovi & Goldberg, 2020). LLM models were complex and opaque, because the deep neural networks that underpinned them were composed of billions of parameters, leading to emergent behaviors that were often unpredictable and not easily interpretable (Lin et al., 2023). Experts further showed that GenAI provided an illusion of transparency, but that what GenAI gave for explanation did not always match what models actually focused on to provide their outputs (e.g., Jain & Wallace, 2019; Dasgupta et al., 2022). Indeed, sometimes models did not actually rely on the chains-of-thought they purported to when reasoning, so relying on these could create a false sense of security in the user (Turpin et al., 2024). Thus, experts recommended avoiding GenAI use where a high degree of explainability was required (Bender et al., 2021; Liu et al., 2023). They also recommended providing the user with global explanations about the model logic and about how to improve the input, since it was not possible to explain the model process for a specific output (Liao and Vaughan, 2023).

6.1.3. Some juniors lacked a deep understanding of GenAI contextualization

Finally, some consultants believed that GenAI was not capable of contextualization. Thus, these juniors recommended only using GenAI for cases where contextualization was not necessary:

Jnr20: "We could implement AI without context in limited areas, where it's fine to have generic answers. So, places like knowledge research, industry trends, slides creating transactional or tactical words that require minimal human touch. Wherever a generic answer is applicable, it's fine for consultants to say, 'I searched using AI.'"

In contrast, articles by GenAI experts available at the time showed that GenAI was strong at contextualization, with the appropriate prompting methods. They noted that effective communication with generative AI required providing contextual information, and specifying the desired output (e.g., Zhou et al., 2023). They also recommended using Retrieval-Augmented Generation (RAG) methods to complement the general knowledge of LLMs with internal data to provide contextually relevant answers (e.g., Gu et al., 2020; Lewis et al., 2020). For example, a consultant could ask a RAG AI agent to provide a summary of internal documents related to a customer organization.

6.2. Novice risk work type #2: Overestimating the potential of changing human routines rather than system design

The second type of novice tactics we discovered was juniors overestimating the potential for managing GenAI output risks related to accuracy, explainability, and contextualization by changing human routines rather than system design. Some examples of this type of risk work are given below, with additional examples shown in **Appendix D, Table D2**.

6.2.1. Juniors recommended change to human routines rather than system design to mitigate accuracy risks

Some consultants suggested managing output risks related to accuracy in two ways. First, they recommended training users to validate results:

Jnr76: "We should have content sessions where we really drill consultants on questions to get second order insights, and check that they really understand, and are not just relying on the technology."

Jnr63: "[It will be important to teach consultants to] do spot checking of the numbers so that they have confidence in how [GenAI] did the approach, and [teach them to do] the critical thinking of, does the approach and does the answer it gave me makes sense. And to think about sources and actually going and digging up each of those sources to make sure they are real, and that is what that source is saying."

Second, they recommended having managers review consultant prompts and responses:

Jnr12: "It would be good to show the manager what you put into GenAI to get that result, because then managers know for sure what you fed to the AI tool before you generated the output. That would help make sure that you didn't get any wrong information, and would give the manager a record of how you got that, to make sure it's not crazy stuff."

Jnr56: "If you did use it, it could be difficult to cross-check what you did. [Providing our managers with] chat transcripts and retracing our steps of analysis could help. As long as you have ways to collaboratively validate [the output], it would be fine."

In contrast, articles by GenAI experts available at the time showed that expecting humans to validate user prompts was very difficult to do, because human users would not always have time to go through activity logs at the speed or scale they desired, or could "fall asleep at the wheel"—fail to exert effort and remain attentive, allowing the AI to substitute, rather than augment their performance (Dell'Acqua et al., 2023).

For this reason, experts suggested managing output risks related to accuracy by making changes to system design. This could be done by fine-tuning a model's parameters based on additional, specialized data (e.g., Devlin et al., 2018; Lee et al., 2019). It could also be done by setting up automatic monitoring with a second system (such as a classifier, or a generative AI system capable of producing its own chains-of-thought; Shavit et al., 2023; Saunders et al., 2022) And it could be done by using models that provided links to sources (e.g., Simkute et al., 2024) or using more accurate models that improved accuracy by combining retrieved data with the generative content from LLMs with meticulous claim-by-claim fact checking (Min et al., 2023).

6.2.2. Juniors recommended change to human routines rather than system design to mitigate complacency risks

Some consultants overestimated the potential of changing human routines, rather than system design, to address the risk of automation complacency in two ways. First, they suggested teaching consultants to take ownership of work done in collaboration with GenAI:

Jnr27: "Consultants need to know that the ownership is theirs, at the end of the day. If they use AI or not, they are presenting this answer; it is on them. The excuse can never be, 'I put this into AI; it's a black box.' They have to understand it. Consultants can't siphon the responsibility off. They own that answer."

Jnr23: "AI is not this invaluable tool; it is still your decision. [Consultants] are not in a position to say, 'AI did this.' They need to learn that it is still their decision, their data they're presenting."

Second, consultants suggested teaching managers to push back on consultants who seemed to be just copying and pasting from GenAI:

Jnr45: "I think there's a feeling sometimes among managers that, 'Oh, if my associate consultant is using generative AI, they're just literally putting everything into GPT and copying and pasting what it says.' So managers need to know how to make sure that consultants have an idea of what it takes to get the best output from generative AI. So that if [managers] do see there is a consultant who is just copying and pasting everything, [the manager] can push them by saying, 'No. You need to get more out of this."

While GenAI experts writing at the time did not disagree with these methods, experts also showed that juniors' focus on changing human routines ignored system design-focused approaches effective for managing complacency risk. For example, experts noted that a GenAI system could be designed to provide proactive self-reflective prompts to help end-users calibrate their confidence in system outputs, asking, "How confident are you in understanding this output? Does anything require explanation?" (Gmeiner et al., 2023). A GenAI system could also be designed with an interface that visualized uncertainty, because highlighting uncertain content could build awareness that AI-generated content could be wrong (Vasconcelos, Bansal, et al., 2023). Further, a GenAI system could support pattern-matching between GenAI suggestions and users' task goals; for example, the system's output could have keywords highlighted, such as variable names or function calls, that would indicate code fit (Barke et al., 2023). Finally, the models' default behavior could apply hard-coded restrictions to improve an LLM's alignment to a custom objective (Lu et al., 2021).

6.2.3. Juniors recommended change to human routines rather than system design to mitigate contextualization risks

Some consultants focused on managing contextualization risks by gaining agreement within the team around using GenAI for documents that did not require a high degree of contextualization

Jnr31: "We just need to get agreements within the team…we can all agree that a 90% version of a document shouldn't be from GPT, because in consulting, we come up with creative solutions. Every client needs a unique solution, and that's hard to do with a GPT template."

Other consultants recommended controlling contextualization risks by training consultants in prompt engineering:

Jnr4: "[We need] training on how to better structure prompts. How to make sure that you are giving cues in terms of tone, audience, things like that."

Jnr15: "We should train people on how to coach GPT to do certain things by teaching them techniques like prompt engineering."

While GenAI experts agreed that training users in prompt engineering was an accepted method for improving contextualization, experts also highlighted system-based approaches for improving contextualization. In addition to using RAG, as described above, experts recommended aligning the model's generations to particular objectives specified by users by using human-labeled preference data at the fine-tuning stage (e.g., Song et al., 2023), designing GenAI systems to begin with a prompt to the user to communicate their goals and preferences to the system (Zamfirescu-Pereira et al., 2023) and creating prompts centrally and building these into the system (e.g., Achiam et al., 2023).

6.3. Novice risk work type #3: Overestimating the potential of intervening at the project level rather than system deployer- or ecosystem-level

The third type of novice tactics we discovered was juniors' overestimating the potential for intervening at the project level (rather than at the system deployer- or ecosystem-level) to manage GenAI output risks related to accuracy, explainability, and contextualization. Some examples of this type of risk work are given below, with additional examples shown in **Appendix D, Table D3**.

6.3.1. Juniors recommended interventions at the project level rather than system deployer- or ecosystem-level to mitigate accuracy risks

Some consultants suggested managing risks related to accuracy at the project level by having consultants and managers agree on the conditions under which GenAI could be used reliably

Jnr44: "Managers need to set expectations explicitly at the start of the case about how GenAI can be used. You're going to run into issues when you don't define those guardrails clearly, and then you'd potentially run into conflicting expectations in the middle of the case that managers probably want to avoid."

Jnr13: "What managers could do is specify what they allow [GenAI] to be used for and how to check its use. Managers need to put safeguards around how they want it used."

Other consultants recommended having managers review consultants' work process around working with GenAI:

Jnr5: “Managers will need to ask consultants, ‘What did you put in [to GPT]? How did you structure your analysis?”

Jnr11: “If I were a manager now, I would set 30 min for each analysis, and tell the consultant to walk me through how they did it with GenAI.”

In contrast, articles by GenAI experts available at the time highlighted that, because LLM developers were often not system deployers, and because LLMs provided increased accessibility of powerful models to users of varying skills and technical backgrounds, it was also important to intervene at the system deployer- and LLM developer-level to mitigate accuracy risk. GenAI experts recommended several actions that could be taken by system deployers to mitigate accuracy risks: a) communicate to users the intended conditions under which GenAI could be used reliably (Liao & Vaughan, 2023); b) provide co-audit tools (Mündler et al., 2023); c) create a prompt library (Svendsen & Garvey, 2023); d) continually assess the alignment of LLMs vis-à-vis evaluation metrics (e.g., Otani et al., 2023); and e) develop evaluation methods that took into account emergent capabilities as models got more capable, and were open ended enough to detect unforeseen risks (Achiam et al., 2023). For example, experts suggested that system deployers could provide users with access to ChatProtect, an AI-based co-audit tool with features to detect and remove hallucinated content from generated text. The co-audit experience let the user inspect different sentences to detect hallucinations via sampling multiple times from the LLM (Mündler et al., 2023). Another approach recommended by GenAI experts was for system deployers to develop robust evaluation frameworks, and create a monitoring role within the organization to continually assess the alignment of LLMs vis-à-vis these evaluation metrics (e.g., Otani et al., 2023).

GenAI experts also recommended intervening at the ecosystem-level to mitigate accuracy risks. Experts highlighted the critical dependency of LLMs on their training data, which, while comprehensive, often failed to encapsulate the rapid changes inherent in real-world contexts, particularly in scenarios post-dating the training period. Thus, experts suggested that LLM developers should assess the representativeness, robustness, and quality of their data sources, and implement mechanisms that allowed LLMs to continually learn from new data, in order to capture recent developments and trends (Li et al., 2023). GenAI experts suggested that developers could do this by creating methods such as dynamically adjusting LLM behaviors by implementing algorithms to assess the recency and applicability of data points (Meng et al., 2021).

GenAI experts further suggested that LLM developers could help to mitigate accuracy risks by: being clear and upfront about how well the GenAI system performed different tasks by explaining its capabilities and limitations (e.g., Solaiman, 2023), by identifying the source materials used to generate it (Liao & Vaughan, 2023), and by flagging and correcting misleading outputs (Ahmad, Tan, Karri, & Pearce, 2023). Further, experts suggested that model developers should refrain from anthropomorphizing the AI, because this could lead people to falsely assume that the AI had greater overall accuracy than it actually did (Vasconcelos, Jörke, et al., 2023).

6.3.2. Juniors recommended interventions at the project level rather than deployer- or ecosystem-level to mitigate complacency risks

Some consultants suggested managing risks related to automation complacency at the project level by having managers give juniors extra time for quality checking

Jnr42: “Frequently we are under time pressure and that could lead to not being able to quality check as much as you’d like. So, ideally, managers would need to give us more time for quality checking. I think that would be important.”

Jnr57: “One challenge is that using GenAI would increase speed of operations leading to increased expectations from managers, but what if [the consultant is] not confident in the output? To mitigate that, we’d need to train managers to give extra time for quality checking.”

Other consultants recommended managers not short selling projects based on expected time savings:

Jnr76: “[Once managers] start thinking that because consultants have [GenAI] they can do things faster, managers will have different expectations…Managers will start to expect this time savings and sell projects based on it. Then, consultants will stop getting details, and will need to compromise on their depth of understanding… Consultants are already working at a high level. We will begin to learn things in one week and think we are experts. If managers don’t give consultants the time, I fear that managers will short sell, and consultants will compromise…Tell managers not to overscope and short sell a project.”

Jnr49: “It could mean that the need to manage [managers’] expectations in terms of output from the team and the number of the people on the team becomes a lot more important.”

While GenAI experts did not disagree with these methods for controlling complacency risk, they also highlighted actions that could be taken at the system deployer- and ecosystem-level to manage risks related to automation complacency. Experts suggested that, at the system deployer-level, to address this risk, LLM deployers could onboard users to the system (De-Arteaga et al. 2020); they could also provide personalized adjustments to users by assessing users’ confidence in their own abilities and adjusting the user experience to help overconfident users develop appropriate reliance (Lu & Yin, 2021). Further, system deployers could design interfaces with cognitive forcing functions that reduced complacency, such as time-outs, on-demand explanations, and asking users to explicitly rule out alternatives (Buçinca et al., 2021). Finally, system deployers could allow users to clearly specify how tasks were allocated between the human and system, to allow users to better distribute the workload according to the respective strengths and weaknesses of humans and GenAI, and to reduce the cognitive demand on users trying to discern the relative responsibilities on a moment-by-moment basis (Simkute et al., 2024).

Experts also recommended interventions at the ecosystem-level to mitigate complacency risks. They noted that GenAI developers could mitigate complacency by building their models to be stringent in rejecting requests that went against their content policy, and by considering the impact on users of the model’s style, tone, or perceived personality (Achiam et al., 2023). Further, they suggested that system feedback could highlight prompt changes and the resulting output changes (Zamfirescu-Pereira et al., 2023). Finally, they noted that developers could support debugging, for example, by providing test cases and test data that users could employ to identify corner cases (Vaithilingam et al., 2022).

7. Discussion

Our findings have implications for both research and practice related to learning around uncertain emerging technologies, risk work in organizations, and human-computer interaction.

7.1. Implications for research

7.1.1. Novice risk work and failed learning around uncertain emerging technologies

We propose the concept of *novice risk work* to highlight that junior professionals may fail to manage risks around uncertain emerging technologies because juniors 1) lack a deep understanding of technologies that have uncertain and wide-ranging capabilities and are changing exponentially, 2) have daily work experience with human-to-human interaction that leads them to focus on change to human routines rather than system design, and 3) have daily work experience at the project level that leads them to focus on interventions at the project level rather than system deployer- or ecosystem- level (Table 1).

Prior literature on technology and learning in occupations assumes that junior professionals are better positioned to learn how to effectively use new technologies than are senior professionals because junior professionals are often closest to the work itself (e.g., Bharatan et al., 2024; Karunakaran, 2018; Kellogg, 2011; Pine & Mazmanian, 2017; Rahman & Barley, 2017), more willing to engage in real-time experimentation with new technologies in front of clients and other occupation members (e.g., Beane, 2019, 2024; Beane and Anthony, 2024; Endacott & Leonardi, 2022; Leonardi, 2007), and open to learning new methods that conflict with existing identities (e.g., Lifshitz, 2018; Nelson & Irwin, 2014), practices (e.g., Beane & Orlikowski, 2015; Mazmanian & Beckman, 2018) and frames (e.g., Anthony & Tripsas, 2016; Mazmanian, 2013).

We show, in contrast, that senior professionals depending on juniors may lead to failed learning around uncertain emerging technologies. In the context of emerging technologies that have a high level of uncertainty in their use, because they have wide-ranging capabilities and are exponentially changing, juniors may lack deep understanding of the technologies' capabilities. In addition, since junior professionals' everyday practice involves human routines, and project-based work, their suggested measures to manage novel risks will likely be shaped by these aspects of their everyday practice. In contrast, experts in emerging technologies such as GenAI may provide different and more useful tactics for managing risk because they better understand the uncertain technologies and have had more experience in their everyday practice with how to best manage risks associated with it.

7.1.2. Novel risks as key barrier to juniors coaching seniors in the use of emerging technologies

Our concept of *novice risk work* also contributes to our understanding of the key barriers to junior professionals being a source of expertise for seniors around the use of new technologies (e.g., Barley, 1986; Kaplan et al., 2017; Kellogg et al., 2021; Leonardi, 2007). The current literature suggests that the key barrier to such coaching is threat to seniors' positional claims associated with juniors overtly coaching seniors. Juniors coaching seniors challenges the historical distinctions of seniors' performance of higher level, more complex tasks (e.g., Barley, 1990; Kellogg, 2009; Leonardi & Bailey, 2008; Van Maanen & Schein, 1977; Zetka, 2003). It challenges seniors' demonstration of expertise in performing these tasks (e.g., Anthony, 2021; Beane, 2019; Beane & Anthony, 2024; Bechky, 2020; Bharatan et al., 2022). And it challenges seniors' enactment of traditional identities, frames, and temporal rhythms while performing the tasks (e.g., Anthony & Tripsas, 2016; Lifshitz, 2018; Nelson & Irwin, 2014; Oborn & Barrett, 2021; Pachidi et al., 2021).

We do not disagree that junior professionals coaching seniors in the use of uncertain emerging technologies may challenge seniors' positional claims. However, we find that, in the case of uncertain emerging technologies, seniors may be more concerned about managing the risks that the technologies pose to seniors' valued outcomes than about managing the preservation of their own positional claims.

Emerging technologies pose novel risks when they 1) have uncertain and wide-ranging capabilities and are changing at an exponential rate, 2) have the potential to surpass human performance in various skilled and cognitive tasks, and 3) combine data and other inputs at an unprecedented scale and detail from a broad ecosystem of actors. Such novel risks may bring threats to seniors' valued outcomes of accuracy of outputs, explainability of outputs, outputs that take into account relevant contextual data, and users' active engagement with and interrogation of outputs.

7.1.3. Novice risk work and risk construction in organizations

Our concept of *novice risk work* also contributes to the understanding of risk construction in organizations. Existing literature on risk construction in organizations focuses on how organizational actors construct "risk objects" by identifying objects deemed to pose risk to values at stake in a particular setting (e.g., Barrett et al., 2024; Corvellec, 2010). This literature highlights that not everything that could be seen as a risk is represented as one; instead, there are often risk construction struggles among differently positioned groups whose different everyday practices lead them to designate objects as risky or not, and to prescribe particular actions to manage risk objects (e.g., Palermo, 2016; Power, 2016a).

We contribute to the understanding of risk construction in organizations in two ways. First, the current literature suggests that actors shape attitudes towards what objects are seen as risk objects and how these risk objects can best be managed when they make risks emotionally salient (e.g., Gale et al., 2016) and use material artifacts to manage the risks (e.g., Power, 2016b). We highlight the important role that junior professionals may play in identifying and managing risk around uncertain emerging technologies, even when they do not make risks emotionally salient or use material artifacts to manage the risks. This is because, in professional organizations, senior professionals often expect junior professionals to coach them in the use of new technologies.

Second, the current literature on risk construction in organizations is agnostic as to whether the existing scientific body of risk knowledge should serve as a basis for organizing risk (Hardy & Maguire, 2020). It is instead focused on how differently positioned groups with different everyday practices designate objects as risky or not, and prescribe particular actions to manage risk objects (e.g., Corvellec, 2010). In contrast, in the context of emerging technologies that have a high level of uncertainty in their use, because they have wide-ranging capabilities and are exponentially changing, we demonstrate that experts who understand the existing scientific body of risk knowledge may provide different and more useful tactics for managing risk around the technologies than junior professionals, because junior professionals may engage in *novice risk work*.

When emerging technologies have uncertain and wide-ranging capabilities that are changing at an exponential rate, junior professionals are likely to be less informed about their capabilities than are experts in the emerging technologies. When emerging technologies have the potential for outperforming humans in a wide variety of skilled and cognitive tasks, junior professionals' focus on change to human routines may be less effective in managing risks than experts' focus changes to system design. And, when emerging technologies depend on a vast, varied, and high volume of data and other inputs from a broad ecosystem of actors, junior professionals' focus on interventions at the project-level may be less likely to be effective than experts' focus on interventions at the system deployer- and ecosystem-level. Thus, In the context of emerging technologies that have a high level of uncertainty in their use, because they have wide-ranging capabilities and are exponentially changing, experts in emerging technologies may provide different and more useful tactics for professionals related to managing risk around the technologies; these experts are likely to better understand the uncertain technologies and to have had more experience in their everyday practice with how to best manage risks associated with them.

However, depending on the risk work of experts around uncertain emerging technologies is not without its challenges. In particular, such experts may be positioned inside of vendor organizations, and this may introduce bias in terms of the types of risk work the experts recommend. For example, such experts may not recommend the need for clarity around training processes and data used to build LLMs. In addition, the experts may fail to recommend the provision of confidence and uncertainty estimations, crucial in high-stakes settings, and may fail to address potential privacy and intellectual property issues (e.g., Gallifant et al., 2024). Even when experts do attempt to engage in such risk work practices, their contextual understanding of a firm or industry may be less expert than that of junior or senior professionals who work in the target setting.

7.1.4. Novice risk work and human-computer interaction around emerging technologies

Finally, our concept of *novice risk work* contributes to our understanding of the human-computer interaction of novice technology users around emerging technologies. Existing literature on human-computer interaction (HCI) highlights that the key limitations to novices' effective human computer interaction are person-specific barriers and technology usability barriers. This literature demonstrates, for example, that novice programmers typically have only a surface knowledge of programs (Ko et al., 2004; Lahtinen et al., 2005; Yang et al., 2018; Zamfirescu-Pereira et al., 2023), and that usability issues may make the technology difficult for novices to learn (Pane & Myers, 1996).

Our concept of novice risk work highlights two different key barriers to novices' effective human-computer interaction around emerging technologies: juniors' social position and novel technology risks. For example, we highlight that, rather than juniors' person-specific barriers, juniors' social position, which may involve the use of human routines in everyday practice and project-based work, can lead novices to recommend managing risk by making change to human routines rather than system design, and by intervening at the project-level rather than system deployer- or ecosystem-level. We further highlight that it is not only the usability of a technology, but also its uncertain and wide-ranging capabilities and exponential speed of change, that may lead novices to lack a deep understanding of the emerging technology's capabilities.

Our elaboration of the importance of juniors' social position and of emerging technologies' uncertain and wide-ranging capabilities and exponential speed of change as critical to novices' human-computer interaction suggests expanding the current HCI literature's focus on addressing person-specific barriers and technology usability barriers in the design of future tools to better support novice users. In particular, the HCI literature could identify interventions related to a broader set of human actors (e.g., system deployers in addition to developers and end-users) and a broader set of material actors (e.g., data and infrastructure in addition to programs, models, and interfaces).

7.1.5. Limitations

Our study has several limitations that could be addressed in future studies of professionals learning to use uncertain emerging technologies. First, the data set comprises interviews with a relatively small set of juniors drawn from one firm, management consultants who are a very particular professional group. Their interests and values and their approach to learning are all quite different from the professional groups studied in other work in the literature on learning to use technologies in occupational work, such as doctors and lawyers. For example, one of the notable characteristics of senior professionals in communities of practice in the classic studies is that many have been in their jobs for 20–30 years. In contrast, many management consultants have a shorter-term horizon and approach tasks with a different perspective. In addition, consultants' sense of what is valuable centers on how they can expedite outcomes important to clients by adding immediate and discernible value, rather than on how they can maintain the quality of a profession in the long term. Because the dynamics of the professional group of consultants are different from groups studied in prior research on communities of practice, future research should investigate whether the dynamics highlighted here are similar or not to those in communities of practice that have been studied in prior literature, and where these dynamics need to be modified or extended.

Second, at the time of the experiment, the technology was still nascent; consultants were in a first exposure moment, so the speed at which the technology was evolving was likely and understandably not apparent to them; juniors might respond differently if they knew that the technology was exponentially changing. The time we gave participants in the field experiment to engage with GenAI was also very limited; it is quite likely that, once given more exposure and training, these same juniors would develop more effective tactics for GenAI use. In addition, all of the interviews were conducted over two months in 2023, so the generalizability of the insights is hampered by the snapshot timeframe. Future research could examine a larger population of professionals, from a wider variety of organizations and industries, after they have spent more time engaging with the technology. Such research could examine which of the challenges we identified persist, and what new challenges emerge.

Third, the emerging technology we studied was GPT-4. As noted, this technology can be easily accessed and customized without coding and without owning infrastructure. Future research could examine the ways in which the challenges and risk work we elaborate are similar and different for other kinds of uncertain emerging technologies that are not as easily accessed and customized.

Fourth, because we studied consultants, but not their managers, we have data on consultant perceptions of managers' potential concerns around consultants' use of AI; we do not have data on managers' actual reactions to consultants' use of AI. In addition, because there was an organizational policy in place at the time of our study that prohibited consultants from using generative AI in their daily work, we have data on consultants' perceptions of the potential risk work that might help to address the novel risks posed by GenAI; we do not have data on risk work practices they actually tried to use. While these are clear limitations of our study, we believe that our study provides an important case, because it offers an early view of both the kinds of novice risk work that junior professionals may suggest for addressing novel risks posed by uncertain emerging technologies, and how these tactics may lead juniors to fail to coach seniors in the use of the emerging technologies.

7.2. Implications for practice

Our findings have practical implications for organizational leaders, developers of emerging technologies, and policymakers.

7.2.1. Implications for organizational leaders

As organizational leaders attempt to keep their employees' skills in sync with uncertain emerging technologies such as artificial intelligence, blockchain, and 3-D printing, they may imagine that juniors will be able to help seniors learn to use the new technologies. In contrast, we highlight that learning to use uncertain emerging technologies may require addressing a novel set of risks and that, in the context of emerging technologies that have a high level of uncertainty in their use because they have wide-ranging capabilities and are exponentially changing, juniors may be novices in addressing such risks. Thus, organizational leaders may need to help junior and senior professionals address emerging technology risks by a) providing a deep understanding of the emerging technology's unique capabilities and limitations, b) making changes to system design in addition to human routines, and c) intervening in sociotechnical ecosystems, rather than only at the project level.

Regarding providing junior and senior professionals with a deep understanding of emerging technology's unique capabilities and limitations, organizational leaders could do this in four primary ways. First, they could provide junior and senior professionals with employee training and awareness programs to explain how the emerging technology works, what the specific risks of the technology are (e.g., risks related to accuracy, explainability, contextualization and automation complacency), and how end users can help to manage these risks. For example, in the case of GPT-4, leaders could provide professionals with training on how to formulate effective prompts, how to interpret the generated outputs, and how to cross-reference outputs using reliable sources and their own domain expertise and knowledge of firm values. Second, leaders could provide junior and senior professionals with access to experts in the emerging technology who could address their questions, provide guidance, and offer best practices for working with the emerging technology. Third, leaders could provide both junior and senior professionals with education in systems thinking, and in the importance of taking system-level approaches for better implementation and use. Fourth, leaders could form a steering group with a mandate for making critical decisions on managing risks associated with the emerging technology. For example, a steering group could have responsibility for determining acceptable use cases for generative AI, and adjusting these targeted use cases according to professionals' learning during AI use and according to newly available models and features.

Regarding making changes to system design in addition to human routines, organizational leaders could do this by identifying specific harms (like accuracy failures) and changing system design to address these harms. For example, at the time of this study, GenAI experts were recommending that accuracy failures could be controlled by setting up automatic monitoring with a second system, by using models that provided links to sources, or by using more accurate models that improved accuracy by combining retrieved data with the generative content from LLMs with meticulous claim-by-claim fact checking. Experts were recommending that contextualization failures could be mitigated by using RAG that aimed at constraining models to rely on a small set of factors and thereby reduce the risk of fabrication of facts.

Finally, regarding intervening in sociotechnical ecosystems, rather than only at the project level, organizational leaders could do this by making changes at the firm-level and ecosystem-level to address particular harms. For example, at the time of this study, GenAI experts were recommending that organizational leaders could take several actions at the firm level to mitigate accuracy risks: a) communicate to users the intended conditions under which GenAI could be used reliably; b) provide co-audit tools; c) create a prompt library; and d) continually assess the alignment of LLMs vis-à-vis evaluation metrics. Experts were recommending that organizational leaders could intervene at the ecosystem level by requiring LLM vendors to assess the representativeness, robustness, and quality of their data sources, and implement mechanisms that allowed LLMs to continually learn from new data in order to capture recent developments and trends. Organizational leaders could also intervene at the ecosystem level by requiring LLM vendors to report on the provenance and curation of the training data, the model's performance metrics, and any incidents and mitigation strategies concerning harmful content.

7.2.2. Implications for developers of emerging technologies

Our paper also has implications for developers of emerging technologies, as it helps to highlight the importance of developers designing emerging technologies that are easier to effectively use by people who are not formally trained in their use. We suggest that developers of emerging technologies may be able to build future tools that better support novice users in the face of their common counterproductive intuitions and struggles.

In the case of generative AI, we highlight that novices may fall into particular traps related to identifying and managing risks related to accuracy, explainability, contextualization, and automation complacency. We suggest that developers could use end-user focused tool design interventions to help minimize such traps. For example, to help reduce automation complacency around GPT-4, developers could design a system that provides the end user with proactive self-reflective prompts, or build an interface that visualizes uncertainty for the end user. Developers could also mitigate automation complacency by considering the impact on users of the model's style, tone, or perceived personality, and by supporting debugging, for example, by providing test cases and test data that users could employ to identify corner cases.

However, our concept of *novice risk work* around uncertain emerging technologies also suggests that novice end users may engage with such technologies in less productive ways not only because of their own novice intuitions, but also because it is extremely difficult to manage risks associated with emerging technologies that have uncertain and wide-ranging capabilities that are changing at an exponential rate, that carry the possibility of outperforming humans in a wide variety of skilled and cognitive tasks, and that combine data and other inputs at an unprecedented scale and detail from a broad ecosystem of actors. Thus, even as we suggest implications for developers of uncertain emerging technologies, we contend that developers' attempts to design their systems to help novices will not be sufficient in addressing the full consequences of these emerging technology risks.

7.2.3. Implications for policymakers

Thus, we suggest that there may be a need to move beyond a focus on the design of human-computer interaction, and focus also on policies that address firm- and ecosystem-levels in which the technologies are developed, deployed, and used. Our focus on end users and our snapshot study in the early period of GenAI implementation does not allow us to generate a comprehensive list of policy recommendations. Here we highlight three key implications for policy that can be derived from our study.

First, it will be important for policymakers to identify risks of harm being generated at multiple levels in the ecosystem—by developers involved in data sourcing and foundation model building and release, by firms involved in model deployment and adaptation, and by end users using the generative capabilities of the models. By identifying risks of harm generated at multiple levels, policymakers can place control mechanisms in place across this network of risk relations to make risk outcomes more understandable and predictable. Second, the nascent, fragmented, and rapidly changing state of generative AI requires policies that are highly responsive to changing needs. Thus, it will be critical for policymakers to examine real world data on the emerging benefits and harms of AI use, and adjust policies in an iterative and cyclical manner (Lee et al., 2022). It will also be critical for policymakers to consider a range of governance approaches, from risk-based approaches (identifying and prioritizing risks in relation to the potential harms AI systems could cause), principles-based approaches (setting out fundamental principles for AI systems, leaving the interpretation and exact details of implementation to organizations), and outcomes-based approaches (focusing on achieving measurable AI-related outcomes without defining specific processes or actions that must be followed for compliance; AIG Alliance, 2024). And it will be necessary for policymakers to cooperate internationally through coordinated multistakeholder efforts, including government, civil society, academia, industry, and impacted communities (AIG Alliance, 2024). Finally, our study highlights the need to engage GenAI experts in policymaking. However, because GenAI experts may be positioned inside of vendor organizations, this may introduce bias in terms of the types of policies the experts recommend. Thus, policymakers will need to assemble a diverse team of differently positioned GenAI experts, and compare their recommendations to those recommended in peer-reviewed articles in the literature on GenAI.

While we suggest this set of practical and theoretical implications, it is early to be making strong claims, particularly given the dynamism of the space at the present time. Forms of GenAI are on the move, and so is our understanding of them. In this paper, we offer preliminary insights suggesting that by identifying interventions related to a broad set of human actors (ecosystem-level developers, firm-level system deployers, and end-user level system users) and material actors (data, models, and infrastructure), we may be better able to manage emerging technology risks related to professionals' valued outcomes.

Funding

We gratefully acknowledge Harvard Business School, MIT Sloan School of Management, Warwick Business School, and The Wharton School for funding this research, and Boston Consulting Group for running the experiment and providing research assistance.

Credit authorship contribution statement

Katherine C. Kellogg: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hila Lifshitz:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Steven Randazzo:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Ethan Mollick:** Writing – review & editing, Writing – original draft, Conceptualization. **Fabrizio Dell'Acqua:** Project administration, Conceptualization. **Edward McFowland:** Conceptualization. **François Candelon:** Writing – review & editing, Project administration, Methodology, Conceptualization. **Karim R. Lakhani:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Data curation, Conceptualization.

Appendix A: Methodological Appendix

A.1. Research setting

We carried out a comprehensive field research study in collaboration with the Boston Consulting Group (BCG), a leading global management consulting firm. This study was designed as a field experiment that aimed to assess how junior professionals—associates and consultants—at the firm used GenAI for two specific tasks: business problem solving and creative product innovation. In this paper, we have concentrated on the consultants who used generative AI (GenAI) for the business problem solving task.

In the business problem solving task, consultants were asked to which identify channels and brands would help a fictional company optimize its revenue and profitability. They were asked to solve this problem using fictitious interview notes with company executives and historical business performance data.

A.2. Data collection

A team of 10 researchers with expertise in qualitative methods interviewed consultants soon after they completed the experiment. We conducted semi-structured, 60-min, formal interviews (Spradley, 1979) via Zoom with associates and consultants who had completed the problem-solving task.

Interviews were conducted to discover both how and when consultants used GenAI to improve their performance on analytical tasks, and their perceptions of the factors and practices that would enable and constrain their future use of GenAI in their daily work. All interviews were recorded (with informants' consent) and transcribed.

A.2.1. Phase 1 data collection

During our first phase of data collection, we interviewed 118 consultants who had completed the problem-solving task with the use of GenAI. Interviews allowed us to deepen our comprehension of both the practices consultants used when collaborating with GenAI during the experiment, and how they expected to use GenAI in their daily work in the future.

In this first phase of interviews, we did not ask any questions about managers, but many of the consultants spontaneously mentioned that they expected that they would need to coach their managers (senior professionals) in the use GenAI. They expected that their managers would perceive particular risks associated with the technology, and that they, as junior professionals, would need to manage these risks in order to coach these senior professionals in the use of GenAI technology.

In an effort to better understand this phenomenon, we re-read in detail the literature on junior professionals coaching senior professionals in the use of new technologies, which highlighted that senior professionals may attempt to learn from junior professionals, who may be better able than their senior counterparts to learn to effectively use the new technologies. We were surprised to find that the literature on juniors coaching seniors did not include a discussion of novel risks that uncertain emerging technologies posed to outcomes that seniors valued nor did it include a discussion of the tactics that the juniors recommended to control the risks that GenAI posed to seniors' valued outcomes. For this reason, we chose to investigate this phenomenon in detail during our second phase of interviews.

A.2.2. Phase 2 data collection

In the second phase of data collection, we interviewed 78 consultants who had completed the problem-solving task with the use of GenAI to gain a deeper understanding of how junior professionals deal with emerging technologies that present novel risks, and what implications this has for occupation members learning to use these technologies. We noted that the tactics the consultants identified appeared to be different from those being recommended by experts in GenAI writing at the time, who had identified similar risks but prescribed different tactics for managing the risks.

A.2.3. Phase 3 data collection

In the third phase of data collection, we searched on google scholar to gather articles from GenAI experts about how to best address this set of risks identified by both junior professionals and GenAI experts—risks to accuracy, explainability, contextualization and automation complacency. We included in our dataset articles publicly posted by GenAI experts before July 2023, so that we could compare tactics suggested by the junior professionals to tactics that had been suggested by experts in GenAI at the same time.

A.3. Data analysis

We employed an inductive analysis approach (e.g., Glaser & Strauss, 1967; Strauss & Corbin, 1990). Our analysis proceeded in several broad phases.

A.3.1. Phase 1 analysis

In the first phase of analysis, we coded our interviews to identify consultants' anticipated potential manager concerns to the use of GenAI in everyday work. We identified that consultants expected that their managers would have concerns about risks related to their valued outcomes of accuracy, explainability, contextualization of results, and automation complacency, and that these risks would limit consultants' ability to coach their managers in the use GenAI, unless consultants could manage these risks.

A.3.2. Phase 2 analysis

In the second phase of analysis, we explored tactics that consultants (junior professionals) reported could mitigate their managers concerns, and so facilitate their ability to coach their managers (senior professionals) in the use GenAI.

A.3.3. Phase 3 analysis

In the third phase of analysis, we compared consultants' suggested tactics to address managers' concerns about the risks that GenAI posed to the valued outcomes of accuracy, explainability, contextualization of results, and automation complacency, to those that were being recommended by experts in GenAI at the same time. We later named consultants' tactics *novice risk work* to delineate tactics that: 1) are grounded in a lack of deep understanding of the emerging technology's capabilities, 2) focus on change to human routines rather than system design, and 3) focus on interventions at the project-level rather than system deployer- or ecosystem-level.

Appendix B. Juniors expected that they would need to educate seniors, but did not expect status threat to be a key barrier

	Additional examples from the interviews with juniors
Juniors expected to educate seniors about effective use of the new technology	**Juniors expected to educate seniors about effective use of the new technology** Jnr21: "I would probably start with them by using AI for the internal portion of our work, [like team communication], so that the managers can get a sense of how things improve with [the use of GenAI]." Jnr38: "You can start by picking the areas that [the managers] do feel comfortable in. For example, like fact aggregation, or whatever, and begin to implement [GenAI] into our ways of working [there], and couple that with learning labs to test it."
No evidence that juniors expected to need to mitigate positional threat to seniors that this upward teaching could raise	**Juniors did not expect to need to teach seniors without appearing to do so, in order to protect seniors' status** Jnr39: "I think making the case to the hesitant manager will go a long way, like, 'You know, it's going to save me time this in this way. And that will give me the opportunity to work on these other aspects of the project that are super important."

Appendix C. Juniors expected the key barrier to upward coaching would be the technology's risks to seniors' valued outcomes

Risks to Valued Outcomes	Additional Examples from Interviews with Juniors
Accuracy	Jnr20: "Managers will be afraid of AI coming up with answers, and junior people will think they know whole world, and will talk to clients which will discount our credibility with them." Jnr32: "I imagine most managers would probably steer us away from using [GenAI], because they don't want to put their reputation at risk if things turn out to be wrong." Jnr13: "[Managers will be concerned that the client] might take some of our results and do stuff with them by hand, and realize that we used generative AI in a way we weren't supposed to, because an Associate was tired at night." Jnr59: "Managers may worry that the more junior people may over-trust GenAI, and provide wrong data or answers to the manager. Then, if the manager doesn't check, they may advise client incorrectly." Jnr60: "I can see my [manager] having an issue with [consultants using GenAI], because they always want the end result to be correct and AI can be wrong."
Explainability	Jnr50: "When [managers] don't understand the tech, they find it hard to relate to it and trust it…We need to decide how consultants should explain the results from AI [to managers]."
Contextualization	Jnr12: "Managers will worry about consultants' lack of context because GPT is very focused on what you just fed it."
Automation Complacency	Jnr13: "What will be difficult for managers…is not micromanaging, and trusting the results provided by especially new team members…[managers will] have a little bit of worry that some people are just ripping into generative AI and passing it along." Jnr5:" I think managers wouldn't want to get into a situation where anyone is overly reliant on AI." Jnr20: "Managers' main concern would be that consultants are not blindly believing everything the tool spits out…there will be people who put zero to 5 % of human touch and rely super heavily on AI." Jnr73: "If I were a [manager], and I learned that the analysis was done using GenAI, I would have a couple of concerns. One is, was there any vetting of the quality of the of the output?" Jnr64: "If there was a lazy AC on the team, and they used AI for everything and didn't bother quality checking, refining it. [Using GenAI] could let them get away with poor quality work. Before AI, you would still have had to come up with something yourself." Jnr70:" I think there's going to be a trust disconnect to a degree, where managers don't believe that the consultants are fully involved in building this thing…consultants may [start] shutting off earlier, and getting to a 90% answer. I think there's going to be a trust issue that I'd very much envision especially if it's [consultants using it for critical] points in the analysis, like if someone's using it [for certain things] like building model outputs."

Appendix D. Additional examples of novice vs. expert risk work tactics

Table D1
Additional Examples of Novice AI Risk Work Tactics Associated with Lack of Understanding.

GenAI capabilities	Novices' suggestions for managing risks reflect misunderstanding of GenAI capabilities	Experts' suggestions for managing risks reflect understanding of GenAI capabilities
Accuracy	**Use a standardized way of asking questions** Jnr63: "Managers could give consultants] a framework of how they should approach [using GenAI], how they should coach the AI tool through that. Then [managers would] be more confident in it, than if consultants just give [GenAI] the question and let it answer it, however, it likes."	• Decide on appropriate use cases where error risks are acceptable (Achiam et al., 2023) • Test GenAI's reliability in executing each subtask (Shavit et al., 2023)
Explainability	**Explain GenAI logic to seniors** Jnr25: "We should…explain its rationale… We have to be able to do a deep dive and explain everything." **Agree on practices for explainable output** Jnr47: "We need to align on what the process is for using [GenAI], and how the output should look. There are some tools that we use a lot…where we are quite aligned on how the output should look. So maybe there needs to be a similar kind of thing with [Gen]AI where there becomes [our organization's] way of using it, and showing it."	• Avoid GenAI use where explainability is required (Bender et al., 2021; Liu et al., 2023) • Provide user with global explanations about model logic and how to improve the input (Liao & Vaughan, 2023)
Contextualization	**Use for cases where contextualization is not necessary** Jnr63: "If you're using it for [a case] that's based off stuff you can search on the Internet, then maybe [using GenAI] makes sense. Like for a project that is sort of generic and in the public domain. I did a project on wildfire policy that probably would have been great to use [GenAI] for, because we spent a lot of time digging up generic information that's available online."	• Provide contextual information, and specify the desired output (Zhou et al., 2023) • Use RAG to add content (Gu et al., 2020; Lewis et al., 2020)

Table D2
Additional Examples of Novice Risk Work Tactics Associated with Change to Human Routines.

Target of Change	Novices Suggest Managing Risks by Changing Human Routines	Experts Suggest Managing Risks by Also Changing GenAI Data, Model, and System Design
Accuracy	**Train users to validate results** Jnr15: "I think managers would struggle if consultants began to rely on GPT too much….And, I can see some consultants copy pasting, too, and getting inaccurate results. To mitigate this, we should train people on how to coach GPT to do certain things." **Have managers review prompts/responses** Jnr7: "Consultants can handle manager concerns by showing managers the prompts they provided to GPT." Jnr26: "Maybe there could be a way of including the prompts you sent AI or some way for managers to check that part…Consultants can say, these are instructions I gave, and managers can tell us if there are any issues with those instructions."	• Set up second automatic monitoring system (Saunders et al., 2022) • Use a more accurate model (Min et al., 2023)
Automation Complacency	**Train users to take ownership** Jnr26: "One big issue is consultants need to learn how to QA the output to take ownership of their work [with GenAI]."	• Build an interface that visualizes uncertainty (Vasconcelos, Jörke, et al., 2023) • Support pattern-matching between GenAI suggestions and users' task goals (Barke et al., 2023) • Apply hard-coded restrictions to improve an LLM's alignment to a custom objective (Lu et al., 2021)
Contextualization	**Train users in prompt engineering** Jnr43: "Give people examples of prompts they can use…what the related outcomes are."	• Align the model's generations to particular objectives specified by users by using human-labeled preference data at the fine-tuning stage (e.g., Song et al., 2023) • Design system to begin with a user prompt to communicate goals (Zamfirescu-Pereira et al., 2023) • Improve prompts centrally and build them into the system (Achiam et al., 2023)

Table D3
Additional Examples of Novice Risk Work Tactics Associated with Project Level Interventions.

Intervention Level	Novices Suggest Managing Risks by Intervening at the Project Level	Experts Suggest Managing Risks by also Intervening at the System Deployer- and Ecosystem -level
Accuracy	**Seniors and juniors on a project agree on conditions under which GenAI can be used reliably** Jnr70: "I imagine a world where a manager would have to set very clear boundaries up front…I think it'll have to be a very clear discussion of, here's my expectation as a manager of when you use it and when you don't. I think it'll be a burden on the manager to set that versus the consultant having to be unsure of when it's okay when it's not okay." Jnr65: "At the beginning of the project, we just need to set the guidelines around what to use it for and not and why." Jnr10: "We'll need to come to agreement within the team around to what degree do we want answers from AI. Should we use it only for first ideas research, or for full analysis." **Managers review consultants' work process** Jnr13: "I think the biggest thing [with GenAI is]…having the manager do what managers currently do, which is sporadically ask questions about how analyses were done. So managers…make sure that the consultant actually did it." Jnr70: "I think it puts an extra level of burden on managers to validate the results. Right now, it's one level [of validation]. A consultant is doing all this work to then validate it…[With GenAI, managers could be concerned] that consultants will start to cut down time on those tasks, because they're using generative AI, and they're doing [only] some level of validation. [Managers may] think that [consultants] are not going to as deeply understand the nuances of the problem, bugs, things that. So, a manager is going to ask questions or try to validate it themselves." Jnr72: "Managers might think that it's not fully our ideas with ChatGPT…[To mitigate this] you could try using it together. Sit down and assess answers together and iterate with it." Jnr77: "Managers will need to be even more skeptical of the results and drilling a bit harder. Because they will wonder about whether or not I checked my work." **Tailor use according to degree of manager acceptance** Jnr38: "I think different managers are going to have different perspectives on how much GPT should be used, and what it should be used for and not. Some will be very supportive of it, others will not… You need to follow their lead…and use it to the degree they feel comfortable. You can begin to leverage it if they do, but if they don't, you shouldn't." Jnr46: "Of course, managers will have their own perception. So, consultants should use it to the degree the manager is comfortable with. And [consultants] should overcommunicate every step." Jnr48: "I would adapt based on my manager's perception of it." Jnr21: "If there is a skeptical [manager], I would probably start using AI for the internal portion of our work, [like team communication], so that the manager can get a sense of how things improve with [the use of GenAI]."	**System Deployer Level** • Provide co-audit tools (Mündler et al., 2023) • Communicate to users the intended conditions under which GenAI can be used reliably (Liao & Vaughan, 2023) • Create a prompt library of effective prompts for particular tasks (e.g., Svendsen & Garvey, 2023) • Develop evaluation methods that take into account emergent capabilities as models get more capable, and are open ended enough to detect unforeseen risks (Achiam et al., 2023) **Ecosystem Level** • Explain system's capabilities and limitations (Solaiman, 2023) • Build in effective explanations (Vasconcelos, Bansal, et al., 2023) • Assess credibility of data sources; use trusted sources (Li et al., 2023) • Adjust LLM behaviors by implementing algorithms to assess the recency and applicability of data points (Meng et al., 2021)
Complacency	**Managers give end users adequate time** Jnr76: "We need to train managers to give consultants enough time to do the work. If my manager does not give me enough time, if I have to choose between doing my work and doing normal human things, I will choose to do human things." **Managers don't shortsell cases** Jnr18: "It will be really necessary to manage [managers'] expectations of what they can receive, if we use Gen AI versus doing it as we've done traditional consultant work."	**System Deployer Level** • Effectively onboard users; Provide personalized adjustments for users (e.g. Lu & Yin, 2021) • Apply cognitive forcing functions (Buçinca et al., 2021) • Allow users to clearly specify human-GenAI task allocation (Simkute et al., 2024) **Ecosystem Level** • Highlight prompt changes and the resulting output changes (Zamfirescu-Pereira et al., 2023) • Support debugging by generating test cases and test data to help identify corner cases (Vaithilingam et al., 2022)

As corresponding author, I declare for my co-authors and me that the submitted work, titled "Novice risk work: How juniors coaching seniors around generative AI and other uncertain emerging technologies can lead to learning failures," was conducted in accordance with the ethical standards of *Information and Organization*. The manuscript is an original piece of work and has not been published previously nor is it under consideration for publication elsewhere.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., … McGrew, B. (2023). *Gpt-4 Technical Report*. arXiv preprint arXiv:2303.08774.

Ahmad, B., Tan, B., Karri, R., & Pearce, H. (2023). *Flag: Finding line anomalies (in code) with generative AI*. arXiv preprint arXiv:2306.12643.

AIG Alliance. (2024). *Briefing Paper Series*. World Economic Forum.

Anthony, C. (2018). To question or accept? How status differences influence responses to new epistemic technologies in knowledge work. *Academy of Management Review, 43*(4), 661–679.

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly, 66*(4), 1173–1212.

Anthony, C., & Tripsas, M. (2016). Organizational identity and innovation. In *, 1. The Oxford Handbook of Organizational Identity* (pp. 417–435).

Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. *Information and Organization, 33*(3), Article 100478.

Bailey, D. E., & Barley, S. R. (2011). Teaching-learning ecologies: Mapping the environment to structure through action. *Organization Science, 22*(1), 262–285.

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., & von Krogh, G. (2022). We are all theorists of technology now: A relational perspective on emerging technology and organizing. *Organization Science, 33*(1), 1–18.

Bailey, D. E., & Leonardi, P. M. (2015). *Technology Choices: Why Occupations Differ in Their Embrace of New Technology*. MIT Press.

Barke, S., James, M. B., & Polikarpova, N. (2023). Grounded copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages, 7*(OOJnrLA1), 85–111.

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, 78–108.

Barley, S. R. (1990). The alignment of technology and structure through roles and networks. *Administrative Science Quarterly*, 61–103.

Barley, S. R. (1996). Technicians in the workplace: Ethnographic evidence for bringing work into organizational studies. *Administrative Science Quarterly*, 404–441.

Barley, S. R., & Kunda, G. (2001). Bringing work back in. *Organization science, 12*(1), 76–95.

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science, 23*(5), 1448–1466.

Barrett, M., Oborn, E., Prince, K., & Lee, E. (2024). *A Relational Perspective on Digital Technologies and Organizing Risk: Telemedicine as a Risk Object in the Re-Imagining of Healthcare Futures* (Working paper).

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

Beane, M. (2024). *The Skill Code: How to Save Human Ability in an Age of Intelligent Machines*. Harper Collins.

Beane, M., & Anthony, C. (2024). Inverted apprenticeship: How senior occupational members develop practical expertise and preserve their position when new technologies arrive. *Organization Science, 35*(2), 405–431.

Beane, M., & Orlikowski, W. J. (2015). What difference does a robot make? The material enactment of distributed coordination. *Organization Science, 26*(6), 1553–1573.

Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization Science, 14*(3), 312–330.

Bechky, B. A. (2020). Evaluative spillovers from technological change: The effects of “DNA envy” on occupational practices in forensic science. *Administrative Science Quarterly, 65*(3), 606–643.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623).

Beunza, D., & Stark, D. (2004). Tools of the trade: The socio-technology of arbitrage in a wall street trading room. *Industrial and Corporate Change, 13*(2), 369–400.

Bharatan, I., Oborn, E., & Swan, J. (2024). Finding your sea legs: Exploring newcomer embodied learning in an extreme context. *Journal of Management Studies, 0*(0), 1–35.

Bharatan, I., Swan, J., & Oborn, E. (2022). Navigating turbulent waters: Crafting learning trajectories in a changing work context. *Human Relations, 75*(6), 1084–1112.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science, 2*(1), 40–57.

Brown, J. S., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science, 12*(2), 198–213.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., … Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early Experiments With gpt-4*. arXiv preprint arXiv:2303.12712.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. In *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1)* (pp. 188:1–188:21).

Cameron, L. D., & Rahman, H. (2022). Expanding the locus of resistance: Understanding the co-constitution of control and resistance in the gig economy. *Organization Science, 33*(1), 38–58.

Campbell, S. (2006). Risk and the subjectivity of preference. *Journal of Risk Research, 9*(03), 225–242.

Chan, C. K., & Hedden, L. N. (2023). The role of discernment and modulation in enacting occupational values: How career advising professionals navigate tensions with clients. *Academy of Management Journal, 66*(1), 276–305.

Corvellec, H. (2010). Organizational risk as it derives from what managers value: A practice-based approach to risk assessment. *Journal of Contingencies & Crisis Management, 18*(3), 145–154.

Dasgupta, I., Lampinen, A. K., Chan, S. C., Creswell, A., Kumaran, D., McClelland, J. L., & Hill, F. (2022). *Language Models Show Human-Like Content Effects on Reasoning*. arXiv preprint arXiv:2207.07051.

De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2020). A case for humans-in-the-loop: decisions in the presence of erroneous algorithmic scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., … Lakhani, K. R. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*. Harvard Business School Technology & Operations Mgt. Unit Working Paper, 24–013.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv: 1810.04805.

Endacott, C. G., & Leonardi, P. M. (2022). Artificial intelligence and impression management: Consequences of autonomous conversational agents communicating on one's behalf. *Human Communication Research, 48*(3), 462–490.

Fayard, A. L., Gkeredakis, E., & Levina, N. (2016). Framing innovation opportunities while staying committed to an organizational epistemic stance. *Information Systems Research, 27*(2), 302–323.

Gale, N. K., Thomas, G. M., Thwaites, R., Greenfield, S., & Brown, P. (2016). Towards a sociology of risk work: A narrative review and synthesis. *Sociology Compass, 10*(11), 1046–1071.

Gallifant, J., Fiske, A., Levites Strekalova, Y. A., Osorio-Valencia, J. S., Parke, R., Mwavu, R., … Pierce, R. (2024). Peer review of GPT-4 technical report and systems card. *PLOS Digital Health, 3*(1), Article e0000417.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*.

Gmeiner, F., Yang, H., Yao, L., Holstein, K., & Martelaro, N. (2023). Exploring challenges and opportunities to support designers in learning to co-create with AI-based manufacturing design tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–20).

Gu, K., et al. (2020). Retrieval augmented language model pre-training. In *International Conference on Machine Learning*. PMLR.

Hardy, C., & Maguire, S. (2020). Organizations, risk translation, and the ecology of risks: The discursive construction of a novel risk. *Academy of Management Journal, 63*(3), 685–716.

Hardy, C., & Thomas, R. (2014). Strategy, discourse and practice: The intensification of power. *Journal of Management Studies, 51*(2), 320–348.

Huising, R. (2015). To hive or to hold? Producing professional authority through scut work. *Administrative Science Quarterly, 60*(2), 263–299.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

Jacovi, A., & Goldberg, Y. (2020). *Towards Faithfully Interpretable NLP Systems: How Should We Define and Evaluate Faithfulness?*. arXiv preprint arXiv:2004.03685.

Jain, S., & Wallace, B. C. (2019). *Attention Is Not Explanation*. arXiv preprint arXiv:1902.10186.

Kaplan, S., Milde, J., & Cowan, R. S. (2017). Symbiont practices in boundary spanning: Bridging the cognitive and political divides in interdisciplinary research. *Academy of Management Journal, 60*(4), 1387–1414.

Karunakaran, A. (2018). *Truce Structures: Examining Cross-Professional Coordination in the Wake of Technological and Institutional Change*. Doctoral dissertation. Massachusetts Institute of Technology.

Karunakaran, A. (2021). In cloud we trust? Co-opting occupational gatekeepers to produce normalized trust in platform-mediated interorganizational relationships. *Organization Science, 33*(3), 1188–1211.

Karunakaran, A., Orlikowski, W. J., & Scott, S. V. (2022). Crowd-based accountability: Examining how social media commentary reconfigures organizational accountability. *Organization Science, 33*(1), 170–193.

Kellogg, K. C. (2009). Operating room: Relational spaces and microinstitutional change in surgery. *American Journal of Sociology, 115*(3), 657–711.

Kellogg, K. C., Myers, J. E., Gainer, L., & Singer, S. J. (2021). Moving violations: Pairing an illegitimate learning hierarchy with trainee status mobility for acquiring new skills when traditional expertise erodes. *Organization Science, 32*(1), 181–209.

Kellogg, K. C. (2011). *Challenging Operations: Medical Reform and Resistance in Surgery*. University of Chicago Press.

Ko, A. J., et al. (2004). Six learning barriers in end-user programming systems. In *Proceedings of the 2004 IEEE Symposium on Visual Languages - Human Centric Computing (USA, 2004)* (pp. 199–206).

Lahtinen, E., Ala-Mutka, K., & Järvinen, H. M. (2005). A study of the difficulties of novice programmers. *ACM SIGCSE Bulletin, 37*(3), 14–18.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lebovitz, S., Levina, N., & Lifshitz, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly, 45*(3).

Lebovitz, S., Lifshitz, 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.

Lee, C., Cho, K., & Kang, W. (2019). *Mixout: Effective Regularization to Finetune Large-Scale Pretrained Language Models*. arXiv preprint arXiv:1909.11299.

Lee, E. L., Barrett, M., Prince, K., Oborn, E., Li, O., & Thomas, P. (2022). *Digital Twins and Service Innovation in Designing the Hospital of the Future*.

Lee, M. Y., Mazmanian, M., & Perlow, L. (2020). Fostering positive relational dynamics: The power of spaces and interaction scripts. *Academy of Management Journal, 63*(1), 96–123.

Leonardi, P. M. (2007). Activating the informational capabilities of information technology for organizational change. *Organization Science, 18*(5), 813–831.

Leonardi, P. M., & Bailey, D. E. (2008). Transformational technologies and the creation of new work practices: Making implicit knowledge explicit in task-based offshoring. *MIS Quarterly*, 411–436.

Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *, 33. Advances in Neural Information Processing Systems* (pp. 9459–9474).

Li, J., Tang, T., Zhao, W. X., Wang, J., Nie, J. Y., & Wen, J. R. (2023). *The Web Can Be Your Oyster for Improving Large Language Models*. arXiv preprint arXiv: 2305.10998.

Liao, Q. V., & Vaughan, J. W. (2023). *AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap*. arXiv preprint arXiv:2306.01941.

Lifshitz, H. (2018). Dismantling knowledge boundaries at NASA: The critical role of professional identity in open innovation. *Administrative Science Quarterly, 63*(4), 746–782.

Lin, Z., Trivedi, S., & Sun, J. (2023). *Generating With Confidence: Uncertainty Quantification for Black-Box Large Language Models*. arXivpreprint arXiv:2305.19187.

Liu, Y., Yao, Y., Ton, J. F., Zhang, X., Cheng, R. G. H., Klochkov, Y., … Li, H. (2023). *Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment*. arXiv preprint arXiv:2308.05374.

Lu, X., Welleck, S., West, P., Jiang, L., Kasai, J., Khashabi, D., … Choi, Y. (2021). *Neurologic a* Esque Decoding: Constrained Text Generation With Lookahead Heuristics*. arXiv preprint arXiv:2112.08726.

Lu, Z., & Yin, M. (2021). Human reliance on machine learning models when performance feedback is limited: heuristics and risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery.

Mazmanian, M. (2013). Avoiding the trap of constant connectivity: When congruent frames allow for heterogeneous practices. *Academy of Management Journal, 56*(5), 1225–1250.

Mazmanian, M., & Beckman, C. M. (2018). "Making" your numbers: Engendering organizational control through a ritual of quantification. *Organization Science, 29*(3), 357–379.

Mazmanian, M., Orlikowski, W. J., & Yates, J. (2013). The autonomy paradox: The implications of mobile email devices for knowledge professionals. *Organization Science, 24*(5), 1337–1357.

Meng, L., Gorbet, R., & Kulić, D. (2021). Memory-based deep reinforcement learning for pomdps. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.

Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W. T., Koh, P. W., … Hajishirzi, H. (2023). *Factscore: Fine-Grained Atomic Evaluation of Factual Precision in Long Form Text Generation*. arXiv preprint arXiv:2305.14251.

Mündler, N., He, J., Jenko, S., & Vechev, M. (2023). *Self-Contradictory Hallucinations of Large Language Models: Evaluation, Detection and Mitigation*. arXiv preprint arXiv:2305.15852.

Nelson, A., Anthony, C., & Tripsas, M. (2023). "If I could turn back time": Occupational dynamics, technology trajectories, and the reemergence of the analog music synthesizer. *Administrative Science Quarterly, 68*(2), 551–599.

Nelson, A. J., & Irwin, J. (2014). "Defining what we do—All over again": Occupational identity, technological change, and the librarian/internet-search relationship. *Academy of Management Journal, 57*(3), 892–928.

Nicolini, D., Pyrko, I., Omidvar, O., & Spanellis, A. (2022). Understanding communities of practice: Taking stock and moving forward. *Academy of Management Annals, 16*(2), 680–718.

Oborn, E., & Barrett, M. (2021). Marching to different drum beats: A temporal perspective on coordinating occupational work. *Organization Science, 32*(2), 376–406.

Orr, J. E. (1990). *Talking About Machines: An Ethnography of a Modern Job*. Cornell University.

Otani, M., Togashi, R., Sawai, Y., Ishigami, R., Nakashima, Y., Rahtu, E., … Satoh, S. I. (2023). Toward verifiable and reproducible human evaluation for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14277–14286).

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science, 32*(1), 18–41.

Palermo, T. (2016). *Technoculture: Risk Reporting and Analysis at a Large Airline*.

Pane, J. F., & Myers, B. A. (1996). *Usability Issues in the Design of Novice Programming Systems* (pp. 96–132). Carnegie-Mellon University; Department of Computer Science.

Pine, K., & Mazmanian, M. (2015). Emerging insights on building infrastructure for data-driven transparency and accountability of organizations. In *iConference 2015 Proceedings*.

Pine, K. H., & Mazmanian, M. (2017). Artful and contorted coordinating: The ramifications of imposing formal logics of task jurisdiction on situated practice. *Academy of Management Journal, 60*(2), 720–742.

Polykarpou, S., Barrett, M., & Oborn, E. (2020). Place and organizing for emerging technologies-challenges of scaling 3DPrinting across a UK Hospital. In *Academy of Management Proceedings (Vol. 2020, no. 1), Briarcliff manor, NY* (p. 20450). Academy of Management.

Power, M. (2016a). *Introduction–Riskwork: The Organizational Life of Risk Management, Riskwork* (pp. 1–25). Essays on the Organizational Life of Risk Management.

Power, M. (2016b). *Postscript: On Riskwork and Auditwork, Riskwork* (pp. 274–284). Essays on the Organizational life of Risk Management.

Pratt, M. G., Rockmann, K. W., & Kaufmann, J. B. (2006). Constructing professional identity: The role of work and identity learning cycles in the customization of identity among medical residents. *Academy of Management Journal, 49*(2), 235–262.

Rahman, H. A., & Barley, S. R. (2017). Situated redesign in creative occupations–an ethnography of architects. *Academy of Management Discoveries, 3*(4), 404–424.

Rahman, H. A., Karunakaran, A., & Cameron, L. D. (2024). Taming platform power: Taking accountability into account in the management of platforms. *Academy of Management Annals, 18*(1), 251–294.

Rahman, H. A., & Valentine, M. A. (2021). How managers maintain control through collaborative repair: Evidence from platform-mediated "gigs.". *Organization Science, 32*(5), 1300–1326.

Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. (2024). *Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies and Limitations*. IEEE Access.

Saunders, W., Yeh, C., Wu, J., Bills, S., Ouyang, L., Ward, J., & Leike, J. (2022). *Self-Critiquing Models for Assisting Human Evaluators*. arXiv preprint arXiv:2206.05802.

Schneider, J., Abraham, R., & Meske, C. (2024). *Governance of Generative Artificial Intelligence for Companies*. arXiv preprint arXiv:2403.08802.

Sergeeva, A. V., Faraj, S., & Huysman, M. (2020). Losing touch: An embodiment perspective on coordination in robotic surgery. *Organization Science, 31*(5), 1248–1271.

Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., … Robinson, D. G. (2023). *Practices for Governing Agentic AI Systems, Research Paper, OpenAI*.

Simkute, A., Luger, E., Evans, M., & Jones, R. (2024). *"It Is There, and You Need It, So Why Do You Not Use It?" Achieving Better Adoption of AI Systems by Domain Experts, in the Case Study of Natural Science Research*. arXiv preprint arXiv:2403.16895.

Solaiman, I. (2023, June). The gradient of generative AI release: Methods and considerations. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 111–122).

Song, F., Yu, B., Li, M., Yu, H., Huang, F., Li, Y., & Wang, H. (2023). *Preference Ranking Optimization for Human Alignment*. arXiv e-prints, arXiv-2306.

Spradley, J. (1979). *The Ethnographic Interview*. Winston, New York: Holt Rinehart &.

Steimers, A., & Schneider, M. (2022). Sources of risk of AI systems. *International Journal of Environmental Research and Public Health, 19*(6), 3641.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research, (Vol. 15).*. Newbury Park, CA: Sage.

Svendsen, A., & Garvey, B. (2023). An outline for an interrogative/prompt library to help improve output quality from generative-AI datasets. In *Prompt Library to Help Improve Output Quality From Generative-AI Datasets (May 2023)*.

Treem, J. W., & Leonardi, P. M. (2017). Recognizing expertise: Factors promoting congruity between individuals' perceptions of their own expertise and the perceptions of their coworkers. *Communication Research, 44*(2), 198–224.

Turpin, M., et al. (2024). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. In *, 36. Advances in Neural Information Processing Systems*. V1 submitted before July 2023 on open archive.

Vaithilingam, P., Zhang, T., & Glassman, E. L. (2022, April). Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–7).

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226.

Van Maanen, J. E., & Schein, E. H. (1977). *Toward a Theory of Organizational Socialization*.

Vasconcelos, H., Bansal, G., Fourney, A., Liao, Q. V., & Vaughan, J. W. (2023). *Generation Probabilities Are Not Enough: Exploring the Effectiveness of Uncertainty Highlighting in AI-Powered Code Completions*. arXiv preprint arXiv:2302.07248.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), 1–38.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science, 33*(1), 59–82.

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour, 7*(9), 1526–1541.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., … Fedus, W. (2022). *Emergent Abilities of Large Language Models*. arXiv preprint arXiv:2206.07682.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., … Gabriel, I. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214–229).

Yang, Q., Suh, J., Chen, N. C., & Ramos, G. (2018). Grounding interactive machine learning tool design in how non-experts actually build models. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 573–584).

Zamfirescu-Pereira, J. D., 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* (pp. 1–21).

Zetka, J. R. (2003). *Surgeons and the Scope*. Cornell University Press.

Zhou, K., Jurafsky, D., & Hashimoto, T. (2023). *Navigating the Grey Area: How Expressions of Uncertainty and Overconfidence Affect Language Models*. arXiv preprint arXiv:2302.13439.