AI Resilient Future of Work: A Counterfactual Lens on Human-AI Collaboration

Vedant Das Swain
New York University
New York, NY, USA
vd2538@nyu.edu

Koustuv Saha
University of Illinois Urbana-Champaign
Urbana, IL, USA
ksaha2@illinois.edu

Abstract

Contemporary discourse on the future of work commonly frames it as an AI-enabled trajectory, with AI expected to augment, automate, and transform everyday labor. While this framing highlights real opportunities, it embeds a strong assumption: that access to AI will be widespread, affordable, and acceptable across workplaces. In this position paper, we challenge this assumption by articulating a counterfactual yet plausible world in which workers lack access to AI due to economic, privacy, legal, or organizational constraints. To examine these risks, we introduce the Counterfactual Resilience Framework (CReF), an analytical lens that treats AI-absence as a diagnostic probe rather than a hypothetical anomaly. Using illustrative workplace scenarios, CReF investigates three lines of inquiry: how AI is unevenly distributed across roles and organizations, how sustained AI-use erodes the skills workers need when AI is withdrawn, and how institutions can mitigate the shock of AI disruption. We conclude that AI-resilience must be developed in parallel with advances in human-AI collaboration—not as an afterthought—to prevent a fractured landscape of AI-enabled and AI-excluded workers.

ACM Reference Format:
Vedant Das Swain and Koustuv Saha. 2026. AI Resilient Future of Work: A Counterfactual Lens on Human-AI Collaboration. In *ACM AI Summit 2026 (AI Summit '26), August 31-September 02, 2026, Atlanta, GA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3806096.3844865

1 Introduction

> An urgent care center adopts a new generative AI tool to automate one of the most tedious aspects of clinical work, "charting," the process in which doctors document their cases and diagnoses for posterity. This newfound productivity is intended to free up resources for doctors to treat more emergencies in their ever-chaotic daily lives of providing care. But the hospital system this ER operates within gets afflicted by an IT outage. Not only is AI out of reach for charting, but they also lose their traditional EMR system for triaging and tracking cases and progress. Lives still need to be saved, so the attendees step up and distribute clipboards to the residents and interns. They go analog and get back to what they do best.

These are incidents depicted in Season 2 of HBO's *The Pitt* [1] that were inspired by real-world technology breakdowns that impact work [49]. The arc illustrates how AI is getting entangled in our work, but also forces us to confront what would happen when these technologies are stripped away. In fact, it even highlights what organizations need to remain *resilient* in the face of AI integration.

The normalization of generative AI promises to augment work in ways that go beyond data synthesis, drafting, and translation. It is becoming a salient part of decision-making, learning, and even hiring. The impact is domain-agnostic and touches white-collar and frontline labor, from generative writing assistants embedded in office suites to algorithmic management systems that mediate scheduling and evaluation in platform and service work [2, 29, 42, 44, 48]. As a consequence, we are witnessing a tide of research focusing on human-AI interaction and collaboration that often centers on how AI technologies should be refined and how people's workflows should be redesigned to maximize benefits. However, as reflected in the vignette above, the push to apply AI risks overlooking what we can do without it.

Yet even as AI becomes embedded in everyday work, permanent access to these tools is far from guaranteed. Digital outages and cyberattacks have already shown how dependence on third-party infrastructure can cascade across work settings and force workers to improvise manual workarounds [32]. AI access may be similarly interrupted not only by technical failure, but also by cost, vendor dependence, privacy constraints, regulatory uncertainty, liability exposure, reputational risk, or institutional bans. Recent federal restrictions on Anthropic systems reveal how access to AI tools can quickly become entangled in legal, political, national-security, and organizational disputes [6]. In other words, the future of work may not simply be AI-augmented; it may also be intermittently AI-absent. While considerable effort is being devoted to preparing workers for AI integration, far less attention has been paid to whether organizations are prepared for moments when AI is unavailable, withdrawn, banned, or rolled back.

Against this backdrop, we adopt a *counterfactual lens* to consider a future of work where AI is not ubiquitously available and human-AI collaboration is not a universal capacity; instead, it is a contingent infrastructural condition that is situational, revocable, and temporary. We reflect on an alternative to the AI-augmented future by considering AI-absent work settings that stem from plausible pathways from today's state of work and AI: 1) AI becomes unaffordable or access is uneven due to tiered enterprise pricing and vendor consolidation; 2) privacy and confidentiality constraints lead organizations (e.g., healthcare, legal, public sector, defense) to prohibit external AI tools; 3) regulatory and liability regimes raise the risk of AI-assisted decisions; 4) organizations restrict AI to reduce reputational risk or to reassert managerial control; or 5) critical incidents, such as data leaks, hallucination-driven errors, or compliance failures, prompt bans or rollbacks.

[1] Season 2, Episode 8 "1:00 P.M." (aired on February 19, 2026).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
AI Summit '26, Atlanta, GA, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2638-5/2026/08
https://doi.org/10.1145/3806096.3844865

Through its analytical lens, this position paper foregrounds three critical concerns about the extant future of work: **(1) distributional questions**—who gets AI and how; **(2) preparedness questions**—how do workers and organizations maintain competence while assimilating AI workflows; **(3) mitigation questions**—can organizations respond to AI-centered shortages? By starting from the possibility of AI-absence, we explore the need to preserve human-workplace capabilities, to protect the social fabric of human workflows, and to contractually defend against an unfair future of work [4, 39, 40, 51]. We describe how the effects of human-AI collaboration are skewed and potentially debilitating. Our position is not to remove AI, but instead to urge a new imperative where AI-resilience is developed in parallel to advancements in human-AI collaboration at work and not as a consequence of it.

2 Tech Coupling, Disruption, and Resilience

Research has long emphasized that technologies at work are not neutral add-ons but sociotechnical arrangements that reorganize practices, relations, and what counts as "good work" [24, 39, 50, 51]. From this perspective, AI is not simply a discrete tool deployment that is disembodied from the user and the environment. Like most digital and algorithmic tools, AI is a *layer of workplace infrastructure* that redistributes attention, discretion, expertise, and responsibility. Human–AI interaction, and by extension collaboration, is not simply a linking of resources but a complex entanglement. It raises problems of calibration, intelligibility, and trust—including how systems should disclose uncertainty, support user control, and manage expectations over time [2, 15, 23, 24, 48, 54].

Furthermore, STS research foregrounds how seemingly automated systems rely on extensive *invisible* human work (e.g., data labeling, moderation, exception handling, and repair), revealing that advanced technology like AI needs to be viewed as a socio-organizational coupling rather than a self-contained artifact [26, 50]. Therefore, the advantages through AI must be viewed by situating its deployment in local practice, how accountability is allocated, and whose values are encoded into "success" criteria [37, 39, 51]. This raises a new question: When a resource (AI) is removed from a system (organization), does it leave the system as it was before it was introduced, or does it create a new floor?

Organizational disruption is inevitable. The public views major disruptions to an organization through externalities, such as a power blackout due to a storm, or perhaps an aggressive contagion spreading through a workforce. However, workers deal with crises every day. It is not uncommon to be in a role where you are proverbially "putting out fires." Historically, to avoid such challenges, many viewed organizations as "machines" that need to be engineered better, until Perrow's groundbreaking sociological perspective in *Normal Accidents* [41]. His findings show that when failures arise, we often blame human error on a worker *operating* the technology, whereas the tightly coupled sociotechnical system of work leaves limited room to buffer. The engineering approach tries to minimize these errors through additional technological scaffolds, but these only increase complexities and do not address the deeper issue that arises from embeddedness. Perrow's work uses nuclear reactors, and particularly the *Three Mile Island* disaster [2], as case studies to explain the fragility of seemingly robust decisions. Recent work on failure modes in multi-agent systems has recognized the need for improved organizational awareness, but in themselves have only scratched the technological aspects [9]. Our paper tries to invert Perrow's approach and other methods of failure analysis. In contrast to retrospective analyses of disruptions, we consider a counterfactual prospective reality. Through this process, we hope to not only uncover where failures might arise, but describe the assumptions underlying the existing systems to answer *distributional questions*.

Understanding failures helps us derive a path to withstanding them. This concept has received attention in organizational literature as a construct called *resilience*. It is an organization's ability to withstand challenges and continue to operate effectively [5]. Moreover, research on resilience reinforces that not all disruptions are acute shocks; many impacts are everyday instances (e.g., staff shortages) and therefore we need to think of "everyday resilience" [5]. Following from the pitfalls highlighted in *Normal Accidents*, we also learn that enforcing resilience top-down through monolithic structures and redundancy is often counterproductive when we overlook empowering and enabling looser interactions between workers and technology. In fact, resilience requires individuals and teams to improvise and self-organize into ad-hoc workflows that maintain performance [55]. We focus on its most human-centered aspects, while acknowledging that other aspects of resilience, such as better organizational tools (e.g., safety audits) remain important. We are specifically motivated by aspects in Duchek's metareview. Resilience requires cognitive capabilities in workers such as mindfulness, sensemaking, and critical thinking [22]. These dimensions help us explore the *preparedness questions*. Apart from preemption, we also need to understand reaction. The literature points to the importance of remaining socially connected, adopting new power structures, and sharing responsibility [56]. We use these dimensions to inquire into the *mitigation questions*.

3 Toward an AI-Resilient Future of Work

The prevailing vision of AI in the workplace emphasizes human-augmentation (not replacement). However, this rhetoric is no longer confined to academic conferences and big tech boardrooms; we are witnessing a push for AI as an economic necessity. Organizations are adopting AI to remain competitive, while workers are urged to become "AI-literate" to maintain employability [58]. After all, we have empirical evidence that working with AI leads to gains in productivity and output quality in certain knowledge work tasks [8]. Having said that, encouraging and enforcing AI rapidly implicitly frames the lack of AI adoption as stagnation or resistance. This narrative sidelines critical questions that help evaluate the health of an organization. We engage in these questions by proposing a *Counterfactual Resilience Framework (CReF)* as an analytical lens.

With CReF, we lean into the possibility of AI non-use and highlight how AI is not a *neutral capability* but an amorphous institutional layer that gets further shaped by existing power dynamics, contracts, and governance. This method treats AI-absence, not as a hypothetical anomaly, but as a *diagnostic probe* that reveals hidden assumptions, dependencies, and limitations. CReF is structured to investigate three lines of inquiry by establishing *counterfactual*

[2]The same power plant is projected to return in 2027 to produce energy for one of the largest AI stakeholders in the industry, *Microsoft*.

anchors. These anchors are composed of a *setting*, *event*, and *aftermath*, grounded in contemporary AI use anecdotes from today's workplace. Below we document illustrative examples to critically reflect on knowledge work through CReF and thus inspire uses in different and more specialized domains.

3.1 Distribution

Setting: Alex's role involves high-stakes relationship management across multiple stakeholder groups. They only use AI to help summarize and format emails. Alex's organization keeps wondering why they still have outstanding tasks, when Sam's data analytics team has increased ticket completion since they integrated AI. As a result, Alex would often be frustrated and demoralized due to the expected norms being established by AI-augmented coworkers.
Event: The growing cost of AI makes the organization hesitant to renew their license and instead negotiate with a new model provider. As the current license is suspended, workers are shut out of AI use until a new contract is resolved.
Aftermath: The impact of AI absence on Alex's role is minimal. Sam wonders if Alex's contribution will become more prominent because of their ability without AI. Alex remains unfazed. Their role was never fully understood, and now they no longer need to make AI work.

Alex's experience above exemplifies how AI cannot sufficiently support the most nuanced aspects of their role. Disconcertingly, its presence reinforces the mechanical perception that organizations have of the role. Even before the advent of generative AI, algorithmic technologies have often been considered monolithic solutions. These are systems optimized for narrow objectives, but without careful situation in practice, they often produce outputs that are contextually inappropriate, overconfident, or inconsistent with organizational norms and professional standards [3, 11, 27, 37]. This body of work suggests that while AI can augment some tasks, a worker's role is often a complex composite of tasks which are arranged and interact in non-deterministic ways. Therefore, questions of misalignment and reliability are not simply about improving model performance or user goal attainment, but about who gets meaningful and useful access to this tool.

Sam's contrasting usage of AI describes uneven distribution of technology benefits and a disruption in social dynamics—raising concerns of power and surveillance. In many workplaces, AI is deployed not only to assist workers but also to monitor them, evaluate them, and manage them at scale, thereby intensifying information asymmetries and shifting discretion away from workers [12, 29, 30, 43, 45] and amplifying concerns about autonomy, consent, and dignity [1, 10, 38]. Note, however, we are not only referring to dedicated AI tools for behavior monitoring, but rather dual-purpose use of AI as an indicator of value. KPIs and OKRs are not new to organizations. Someone like Sam might have found a metric like *lines of code* to be too limited in exhibiting how much their team thinks. Perhaps appropriating AI token-use helps make their work visible. Yet, for Alex, "token-maxxing" is just another form of misappropriating limited quantified measurements as productivity goals [45]. One might argue that token-use might get disregarded over time, like lines of code. The challenge is the veneer of generative AI being a general-purpose tool, which assumes that it is robust enough that workers like Alex, no matter how specialized or fluid their role, should be able to demonstrate their work through it. When we start thinking of AI being stripped away from work, it can act as a pressure test of the monitoring infrastructure within an organization, whether that be through data capture or social appraisal.

Alex wanted their work to be rewarded; the presence of AI was making them work faster but it was not fulfilling because of the organizational view of their role. A worker's purpose and what they are assigned to do often lead to challenges of fit and ambiguity [14, 47]. And the existing method of deploying AI tools stands to worsen this. It can lead to a systematic trivialization of certain tasks and, even worse, a systematic reduction of certain roles. Several workers like Alex are compelled to fit a square peg into a round hole and spend productive time trying to make AI work. In fact, AI often externalizes labor into hidden supply chains of data work and moderation, raising questions about whose work is made visible, valued, and protected [16, 20, 26, 50]. This thread of inquiry forces us to reflect on the political economy of automation and motivates the need to rethink our view of work. Recent literature shows that while excessive AI adoption might obfuscate intense forms of traditionally feminized labor (e.g., data work), it could recenter the importance of other forms of *feminized* labor (e.g., emotion work) [20, 31, 33]. Questioning the politics of distribution helps reframe access from an issue of accommodating edge cases to an issue of developing core cultural changes at work.

Takeaways from CReF: Organizations are compromised when their evaluation implicitly assumes homogeneous AI use. Performance is expected without a sense of the jagged frontier of AI [21] and without pluralistic models of performance [11, 12, 18]. Employment contracts, performance metrics, and workplace norms increasingly embed assumptions about AI availability. In some settings, AI use is mandatory; in others, it is prohibited; in still others, it is unevenly distributed across roles and seniority. By conceiving a counterfactual anchor, organizations can identify what these settings and scenarios of uneven AI-use are to ensure that evaluation of work adapts to this jagged frontier of AI. In doing so, CReF can steer new forms of accountability, evaluation, and fairness.

3.2 Preparedness

Setting: Tanya is a junior compliance analyst who has spent the past two years reviewing AI-drafted risk assessments, checking them for accuracy before they reach her manager, Max, who still drafts everything by hand out of habit. Tanya rarely drafts assessments from scratch anymore; the AI handles the first pass, and she can overcome writing blocks. But Max often jokes that Tanya's generation "grew up with training wheels."
Event: A regulatory audit flags inconsistencies in severe threats in the AI vendor's model deployment. The organization suspends AI use across the compliance team to protect confidential data. With no timeline for reinstatement, analysts are told to resume manual drafting immediately.
Aftermath: Max lobbies to relax performance expectations for the quarter citing disruption and transition of tools among her direct reports. Despite this protection, Tanya struggles to engage and immerse without the instruments she is familiar with, even though compliance is her domain. If only she could supervise someone to do the grunt work.

Legal liabilities forced Tanya and Max to forgo AI-assisted work [52]. By choice or force, the restricted AI usage can dampen work performance. This whiplash effect is not that different from the blanket policies organizations employed after COVID-19 restrictions were

lifted, leading to quiet quitting when workers felt their wellbeing was no longer cared for [25]. Unlike the advancements in remote work, the AI-augmented future of work is different in its affordances, which risk asymptotically *de-skilling* workers.

Max's joke might appear as a generational divide, but it echoes longstanding concerns. Scholarship on automation has long cautioned that removing operators from hands-on practice while leaving them responsible for rare, high-stakes failures erodes expertise precisely when it is most needed [4]. Similarly, patterns of misuse and disuse, over- and under-reliance, complacency, and reduced situational awareness [40] remain ominous. The very features that make AI valuable in daily routines—speed, fluency, convenience—can also reduce opportunities for deliberate practice, reflective reasoning, and the development of tacit knowledge [23, 35]. This creates a paradox: AI may improve short-term productivity by reducing friction; however, that very friction was providing the footholds for workers to nurture long-term competence.

Max also tried to buffer organizational pressure; however, this possibly revealed a way forward. An AI-resilient workplace is one that anticipates non-linear adoption—in this case, through reworking of social expectations in the face of technological challenges. This approach recognizes that human capacities cannot be easily reconstituted once lost. Workers become increasingly fluent in supervising AI outputs without retaining the capacity for independent reasoning or judgment. As a result, organizations need to rethink how they want to use the time reclaimed from AI efficiency. Most organizations are redirecting that time to more work, but they should consider redirecting it to learning and preservation.

Takeaways from CReF: Within contemporary AI-mediated work, expertise decays gradually and invisibly as AI handles routine reasoning, drafting, and decision support. Workers may remain capable of supervising AI outputs without retaining the ability to independently perform or justify the underlying task. This creates a fragile form of competence that is entangled with AI availability. CReF treats skill erosion as a preparedness risk rather than an individual failure. It asks whether organizations have preserved opportunities for deliberate practice, reflection, and independent reasoning, or whether AI fluency has displaced deeper expertise.

3.3 Mitigation

Setting: Camila is a care coordinator at a residential elder care facility that recently adopted an AI tool to manage resident care plans, flag medication conflicts, and schedule staff rotations. Camila initially welcomed the overdue modernization to replace her usual navigation of the patchwork of spreadsheets and handwritten shift logs. Although eventually she found that AI replaced a fractured workflow, it did not absorb the workarounds, tacit understanding, and institutional knowledge. Coordinators like her were the connective tissue that needed to constantly nudge AI's context.
Event: A reorganization of state funds deprioritizes resources to health centers. The facility is struggling to maintain staff while also affording advanced digital technology.
Aftermath: Camila's team attempts to decouple from the AI workflow by trying to reconstruct their pre-AI tool use so that they can make a case that the AI workflow can be abandoned to preserve staff. However, the organizational memory is no longer scattered in offices; it is hidden in hierarchical files and embeddings that her team cannot decipher. She feels out of her depth as she loses time from casework. Contracting a service team for reverse transition is out of their budget.

Camila's need to embrace decoupling is not a regression of technological advancement but emerges from it as a realistic possibility that has historical precedent. Non-linear technological growth is not uncommon, and we use references of prior technological changes to extrapolate mitigation for AI.

At first, email was celebrated as a frictionless communication medium, but it quickly became a source of overload, inefficiency, and burnout. Over time, organizations developed norms, etiquette, and technical interventions—such as filtering, asynchronous expectations, and alternative communication channels—to mitigate their excess and absence [13, 36, 46, 57]. Similarly, expansive deployment of AI needs a deliberate recalibration at a sociocultural level to protect resources necessary in adverse events.

In the early days of big data, companies aggressively pursued large-scale data accumulation and processing under the assumption that more data would inevitably translate into value. Over time, however, growing awareness of privacy harms, public backlash, and regulatory intervention—such as data protection laws—disrupted this trajectory [7, 59]. Organizations were forced to reconsider the legitimacy and sustainability of data-driven practices and adopt alternative business models. Big data has not vanished, but it has been reconfigured under legal, ethical, and social constraints [37]. Similarly, AI-driven practices are a new monoculture, but organizations need to think about how they recover when their value declines or is in tension with other resources.

Finally, consider the normalization of remote work during the COVID-19 pandemic. Some narratives had positioned this as the inevitable new future of work [13, 17, 19]. However, instead of a linear shift toward fully remote arrangements, many organizations converged on hybrid models that took stock of trade-offs between flexibility, coordination, accountability, and organizational culture [28, 53]. Even with the resurgence of in-person work, remote and hybrid work paradigms have changed how we collaborate, communicate, and allocate our time and resources. Much like these cases, the mitigation for limited AI-use would rely on experimentation with technology itself and on renegotiation of organizational practices.

Takeaways from CReF: AI is valuable for several routine tasks (e.g., drafting, summarization, triage, prioritization, and sensemaking) [34], but it is also creating a new single point of failure. Any degradation in AI use risks stalling work and requires emergency improvisation. However, as we have learned from previous advancements, these improvisations are not immediate. In poorly prepared systems, the burden falls on individual workers, who must compensate through unpaid labor, increased risk, or degraded outcomes. In more resilient systems, institutions absorb the shock through redundancy, adjusted expectations, and protective policies. CReF urges organizations to take stock of both structural and cultural factors to rethink dependency on AI.

4 Conclusion

A future in which workers are limited from collaborating with AI risks deepening inequality and eroding organizational cohesion. Cost, privacy concerns, and compliance constraints can disproportionately exclude all kinds of workers, be it public-sector employees, regulated professions, or contract workers. CReF is a conceptual approach to engage seriously with the counterfactual world of no

AI at work. It is neither a predictive tool nor an evaluative one. It is a generative analysis tool that helps organizations diagnose their human-AI collaboration practices by questioning their readiness when AI becomes scarce. Organizations need to develop *counterfactual anchors* to ask questions along three aspects of resilience: *distribution*, *preparedness*, and *mitigation*. Admittedly, we illustrated CReF through fictional examples, but future researchers can devise rich anchors through participatory research. These anchors are probes that elicit hidden dependencies, contractual obligations, and even potential safeguards in the sociotechnical setting of work. As a conceptual framework, it helps hypothesize more sustainable and healthy integration of AI. This method can help strengthen institutions, protect workers, and ensure that the benefits of AI do not accrue only to those who can afford, tolerate, or accept its costs.

Acknowledgments

This work was partly supported by the Center for Social & Behavioral Studies (CSBS) Small Grant Program at the University of Illinois Urbana-Champaign.

References

[1] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. 2017. Limitless worker surveillance. *California Law Review* 105, 3 (2017), 735–776. doi:10.15779/Z38BR8MF94

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 1–13. doi:10.1145/3290605.3300233

[3] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. *arXiv preprint arXiv:1606.06565* (2016). doi:10.48550/arXiv.1606.06565

[4] Lisanne Bainbridge. 1983. Ironies of Automation. In *Analysis, Design and Evaluation of Man–Machine Systems*. Elsevier, 129–135. doi:10.1016/b978-0-08-029348-6.50026-9

[5] Edwine Barasa, Rahab Mbau, and Lucy Gilson. 2018. What Is Resilience and How Can It Be Nurtured? A Systematic Review of Empirical Literature on Organizational Resilience. *International Journal of Health Policy and Management* 7, 6 (Feb. 2018), 491–503. doi:10.15171/ijhpm.2018.06

[6] Atilla Arda Beşen. 2026. Artificial Intelligence and Communication Channels: Will Great Powers or Corporations Prevail? *Journal of Geography, Politics and Society* 16, 1 (2026), 60–66. doi:10.26881/jpgs.2026.1.06

[7] danah boyd and Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15, 5 (2012), 662–679. doi:10.1080/1369118x.2012.678878

[8] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. 2025. Generative AI at Work. *The Quarterly Journal of Economics* 140, 2 (2025), 889–942. doi:10.1093/qje/qjae044

[9] Mert Cemri, Melissa Z. Pan, Shuyi Yang, Lakshya A. Agrawal, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, Matei A. Zaharia, Joseph Gonzalez, and Ion Stoica. 2025. Why Do Multi-Agent LLM Systems Fail?. In *Advances in Neural Information Processing Systems 38 (NeurIPS '25)*. Neural Information Processing Systems Foundation, 135676–135729. doi:10.52202/085713-4082

[10] Shreya Chowdhary, Anna Kawakami, Mary L Gray, Jina Suh, Alexandra Olteanu, and Koustuv Saha. 2023. Can Workers Meaningfully Consent to Workplace Wellbeing Technologies?. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. ACM, 569–582. doi:10.1145/3593013.3594023

[11] Vedant Das Swain, Lan Gao, Abhirup Mondal, Gregory D Abowd, and Munmun De Choudhury. 2024. Sensible and Sensitive AI for Worker Wellbeing: Factors that Inform Adoption and Resistance for Information Workers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, 1–30. doi:10.1145/3613904.3642716

[12] Vedant Das Swain, Lan Gao, William A Wood, Srikruthi C Matli, Gregory D Abowd, and Munmun De Choudhury. 2023. Algorithmic Power or Punishment: Information Worker Perspectives on Passive Sensing Enabled AI Phenotyping of Performance and Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 1–17. doi:10.1145/3544548.3581376

[13] Vedant Das Swain, Javier Hernandez, Brian Houck, Koustuv Saha, Jina Suh, Ahad Chaudhry, Tenny Cho, Wendy Guo, Shamsi T Iqbal, and Mary Czerwinski. 2023. Focused Time Saves Nine: Evaluating Computer-Assisted Protected Time for Hybrid Information Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 1–18. doi:10.1145/3544548.3581326

[14] Vedant Das Swain, Manikanta D Reddy, Kari Anne Nies, Louis Tay, Munmun De Choudhury, and Gregory D Abowd. 2019. Birds of a Feather Clock Together: A Study of Person-Organization Fit Through Latent Activity Routines. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30. doi:10.1145/3359267

[15] Vedant Das Swain and Koustuv Saha. 2024. Teacher, Trainer, Counsel, Spy: How Generative AI can Bridge or Widen the Gaps in Worker-Centric Digital Phenotyping of Wellbeing. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3663384.3663401

[16] Vedant Das Swain and Koustuv Saha. 2024. Teacher, Trainer, Counsel, Spy: How Generative AI Can Bridge or Widen the Gaps in Worker-Centric Digital Phenotyping of Wellbeing. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. ACM, 1–13. doi:10.1145/3663384.3663401

[17] Vedant Das Swain, Koustuv Saha, Gregory D Abowd, and Munmun De Choudhury. 2020. Social Media and Ubiquitous Technologies for Remote Worker Wellbeing and Productivity in a Post-Pandemic World. In *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 121–130. doi:10.1109/cogmi50398.2020.00025

[18] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M. Gregg, Suwen Lin, Gonzalo J. Martinez, Stephen M. Mattingly, Shayan Mirjafari, Raghu Mulukutla, Subigya Nepal, Kari Anne Nies, Manikanta D. Reddy, Pablo Robles-Granda, Andrew T. Campbell, Nitesh V. Chawla, Sidney D'Mello, Anind K. Dey, Kaifeng Jiang, Qiang Liu, Gloria Mark, Edward Moskal, Aaron Striegel, Louis Tay, Gregory D. Abowd, and Munmun De Choudhury. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4, Article 130 (2019), 27 pages. doi:10.1145/3369828

[19] Vedant Das Swain, Shane Williams, Adam Fourney, and Shamsi T. Iqbal. 2022. Two Birds with One Phone: The Role of Mobile Use in the Daily Practices of Remote Information Work. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '22)*. ACM, 1–8. doi:10.1145/3533406.3533416

[20] Vedant Das Swain, Qiuyue "Joy" Zhong, Jash Rajesh Parekh, Yechan Jeon, Roy Zimmerman, Mary Czerwinski, Jina Suh, Varun Mishra, Koustuv Saha, and Javier Hernandez. 2025. AI on My Shoulder: Supporting Emotional Labor in Front-Office Roles with an LLM-based Empathetic Coworker. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, 1–29. doi:10.1145/3706598.3713705

[21] Fabrizio Dell'Acqua, Edward McFowland, Ethan Mollick, Hila Lifshitz, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2026. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. *Organization Science* 37, 2 (March 2026), 403–423. doi:10.1287/orsc.2025.21838

[22] Stephanie Duchek. 2020. Organizational resilience: a capability-based conceptualization. *Business Research* 13, 1 (April 2020), 215–246. doi:10.1007/s40685-019-0085-7

[23] Upol Ehsan, Samir Passi, Koustuv Saha, Todd McNutt, Mark O Riedl, and Sara Alcorn. 2026. From Future of Work to Future of Workers: Addressing Asymptomatic AI Harms to Foster Dignified Human-AI Interaction. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, 1–21. doi:10.1145/3772318.3791081

[24] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32. doi:10.1145/3579467

[25] Gallup. 2023. Why the World Can't Quit Quiet Quitting. https://www.gallup.com/workplace/507650/why-world-quit-quiet-quitting.aspx Section: Workplace.

[26] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business.

[27] Kaely Hall, Marcus Ma, Xinyue Zhang, Vedant Das Swain, and Jennifer G. Kim. 2025. Understanding Human-AI Misalignment in LLM-Based Job-Seeking Support for Neurodivergent Users. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. ACM, 1–13. doi:10.1145/3663547.3746361

[28] Javier Hernandez, Vedant Das Swain, Jina Suh, Daniel McDuff, Judith Amores, Gonzalo Ramos, Kael Rowan, Brian Houck, Shamsi Iqbal, and Mary Czerwinski. 2025. Triple Peak Day: Work Rhythms of Software Developers in Hybrid Work. *IEEE Transactions on Software Engineering* 51, 2 (2025), 344–354. doi:10.1109/TSE.2024.3504831

[29] Mohammad Hossein Jarrahi, Gemma Newlands, Min Kyung Lee, Christine T Wolf, Eliscia Kinder, and Will Sutherland. 2021. Algorithmic Management in a

Work Context. *Big Data & Society* 8, 2 (2021), 20539517211020332. doi:10.1177/20539517211020332

[30] Anna Kawakami, Shreya Chowdhary, Shamsi T Iqbal, Q Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. 2023. Sensing Wellbeing in the Workplace, Why and For Whom? Envisioning Impacts with Organizational Stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33. doi:10.1145/3610207

[31] Anna Kawakami, Jordan Taylor, Sarah Fox, Haiyi Zhu, and Kenneth Holstein. 2026. AI failure loops in devalued work: The confluence of overconfidence in AI and underconfidence in worker expertise. *Big Data & Society* 13, 1 (March 2026), 20539517261424164. doi:10.1177/20539517261424164

[32] Ethan Larsen, Daniel Hoffman, Carlos Rivera, Brian M Kleiner, Christian Wernz, and Raj M Ratwani. 2019. Continuing Patient Care during Electronic Health Record Downtime. *Applied Clinical Informatics* 10, 03 (2019), 495–504. doi:10.1055/s-0039-1692678

[33] Duri Lee, Heejeong Lim, Vedant Das Swain, and Uichin Lee. 2026. Understanding Behind the Smile of Emotion Workers: Detecting After-Call Stress in Call Agents. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, 1–21. doi:10.1145/3772318.3790803

[34] Weijiang Li, Yinmeng Lai, Sandeep Soni, and Koustuv Saha. 2025. Emails by LLMs: A Comparison of Language in AI-Generated and Human-Written Emails. In *Proceedings of the 17th ACM Web Science Conference 2025 (WebSci '25)*. ACM, 391–403. doi:10.1145/3717867.3717872

[35] Yueshen Li, Hanyi Min, Vedant Das Swain, and Koustuv Saha. 2026. TUX: Measuring Human–AI Tacit Understanding. In *Findings of the Association for Computational Linguistics: EMNLP 2026*.

[36] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. Email Duration, Batching and Self-interruption: Patterns of Email Use on Productivity and Stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 1717–1728. doi:10.1145/2858036.2858262

[37] Brent Mittelstadt. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1, 11 (2019), 501–507. doi:10.1038/s42256-019-0114-4

[38] Phoebe Moore and Andrew Robinson. 2016. The quantified self: What counts in the neoliberal workplace. *New Media and Society* 18, 11 (2016), 2774–2792. doi:10.1177/1461444815604328

[39] Wanda J Orlikowski. 2007. Sociomaterial Practices: Exploring Technology at Work. *Organization Studies* 28, 9 (2007), 1435–1448. doi:10.1177/0170840607081138

[40] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253. doi:10.1518/001872097778543886

[41] Charles Perrow. 2011. Normal Accidents : Living with High Risk Technologies - Updated Edition. (2011), 1–464. https://www.torrossa.com/en/resources/an/5576535

[42] Alex Rosenblat, Tamara Kneese, and <!>danah <!>boyd. 2014. Workplace Surveillance. *SSRN Electronic Journal* (2014), 80–82. doi:10.2139/ssrn.2536605

[43] Alex Rosenblat and Luke Stark. 2016. Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International journal of communication* 10 (2016), 27.

[44] Koustuv Saha and Eshwar Chandrasekharan. 2023. Future of Workplace Conversational AI-Driven Personal Assistants: Promises and Perils. In *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE '23)*. ACM, 133–135. doi:10.1145/3632754.3632939

[45] Koustuv Saha, Pranshu Gupta, Gloria Mark, Emre Kiciman, and Munmun De Choudhury. 2024. Observer Effect in Social Media Use. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, 1–20. doi:10.1145/3613904.3642078

[46] Koustuv Saha and Shamsi T Iqbal. 2023. Focus Time for Wellbeing and Work Engagement of Information Workers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. ACM, 1–11. doi:10.1145/3544549.3585688

[47] Koustuv Saha, Manikanta D Reddy, Stephen Mattingly, Edward Moskal, Anusha Sirigiri, and Munmun De Choudhury. 2019. LibRA: On LinkedIn Based Role Ambiguity and Its Relationship with Wellbeing and Job Performance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30. doi:10.1145/3359239

[48] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504. doi:10.1080/10447318.2020.1741118

[49] Zachary Siegel. 2026. 'The Pitt' Finally Offers a Stark Warning on AI in Medicine. *Wired* (Feb. 2026). https://www.wired.com/story/the-pitt-finally-offers-a-stark-warning-on-ai-in-medicine/ Section: tags.

[50] Susan Leigh Star. 1999. The Ethnography of Infrastructure. *American Behavioral Scientist* 43, 3 (1999), 377–391. doi:10.1177/00027649921955326

[51] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

[52] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. doi:10.48550/arXiv.1711.00399 arXiv:1711.00399 [cs.AI].

[53] Lena Waizenegger, Brad McKenna, Wenjie Cai, and Taino Bendz. 2020. An Affordance Perspective of Team Collaboration and Enforced Working from Home during COVID-19. *European Journal of Information Systems* 29, 4 (July 2020), 429–442. doi:10.1080/0960085X.2020.1800417

[54] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3411764.3445645

[55] Karl E. Weick. 1988. Enacted Sensemaking in Crisis Situations. *Journal of Management Studies* 25, 4 (1988), 305–317. doi:10.1111/j.1467-6486.1988.tb00039.x

[56] Karl E Weick. 1988. Enacted sensemaking in crisis situations. *Journal of management studies* 25, 4 (1988), 305–317.

[57] Steve Whittaker and Candace Sidner. 1996. Email Overload: Exploring Personal Information Management of Email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM, 276–283. doi:10.1145/238386.238530

[58] World Economic Forum. 2023. The Future of Jobs Report 2023. https://www.weforum.org/publications/the-future-of-jobs-report-2023/in-full/

[59] Shoshana Zuboff. 2019. Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum* 28, 1 (Jan. 2019), 10–29. doi:10.1177/1095796018819461