AI, Skill Drift, and Practice Protocols: A Longitudinal Field Study in AI-Augmented Work

Akshaya Jayasankar

Abstract

AI copilots can deliver large short-run productivity gains, but we know little about their long-run effects on workers' underlying skills and career trajectories. This project proposes a longitudinal field study of AI-augmented knowledge work that compares two use profiles—an "autopilot" pattern where workers lean heavily on AI suggestions, and a "sparring partner" pattern where AI is used more selectively as a critique and ideation tool. I additionally introduce a temporal practice protocol in which some teams adopt scheduled "AI-off" windows for unaided work and structured reflection, while others continue with business-as-usual use.

I will instrument fine-grained logs from AI copilots and collaboration tools, link them to quarterly unaided benchmark tasks and to administrative data on staffing, promotion, and high-visibility assignments, and estimate treatment effects using modern difference-in-differences and event-study estimators leveraged on staggered rollout of the practice protocol. This design will allow me to characterize patterns of skill drift versus skill lift, identify who gains or loses expertise and authority under different AI use norms, and test whether deliberate practice preserves expertise without sacrificing efficiency.

By connecting AI adoption to expertise, authority, and inequality inside organizations, the project aims to advance theory on *algorithmic authority* and *practice design* in AI-augmented work, while offering actionable guidance for managers deciding how to govern AI use, evaluation, and promotion on their teams.

1 Introduction

Generative AI tools are rapidly diffusing into knowledge work, promising large jumps in short-run productivity. Field experiments in customer support and professional writing show that AI assistance can both raise average quality and compress performance gaps within tasks [1, 2]. Yet organizations are built not only on short-run output, but also on deep expertise, mentorship, and legitimate authority. It remains unclear whether heavy reliance on AI copilots will produce *skill lift*—accelerating learning and broadening access to expertise—or *skill drift*—subtly eroding workers' unaided capabilities and narrowing future career options.

At the same time, research on algorithms at work shows that new technologies can reshape control, accountability, and inequality inside organizations [3–6]. Existing empirical studies typically observe performance with AI turned on, rather than directly measuring unaided performance trajectories and linking them to authority, visibility, and advancement. Little is known about how day-to-day patterns of AI use interact with practice structure to shape skill, expertise recognition, and opportunity.

This project asks: **Does sustained AI assistance create skill drift in knowledge work, and can a deliberately designed practice protocol preserve both skill and opportunity?** I focus on three intertwined outcomes: (1) workers' unaided task performance over time, (2) who is recognized as an expert and assigned to high-stakes work, and (3) how these patterns differ by role and background. Empirically, I propose a longitudinal field study in AI-augmented teams that combines instrumented logs of AI copilot use, quarterly unaided benchmark tasks, and administrative data on staffing and promotion. The core idea is that *how* people practice with AI—including structured periods of working without it—may be as important as the AI tools themselves.

1.1 Specific Research Sub-Questions

To make this question empirically tractable, I propose the following sub-questions:

- **RQ1 (Skill trajectories).** How do different AI use profiles (e.g., autopilot versus sparring-partner use) relate to changes in workers' unaided performance over 9–12 months?
- **RQ2 (Practice protocol).** Does a temporal practice protocol with periodic "AI-off" windows and structured reflection attenuate skill drift and foster skill lift, relative to business-as-usual AI use?
- **RQ3 (Authority and visibility).** How do AI use profiles and practice protocols affect who is recognized as an expert, who gets staffed on high-visibility assignments, and early career outcomes such as promotions or role transitions?
- **RQ4 (Heterogeneity and inequality).** How do these effects vary by seniority, team, and worker background (e.g., prior experience, gender, or educational pathway), and do they amplify or mitigate existing inequalities?
- **RQ5 (Mechanisms).** How do workers and managers narrate their own learning, ownership, and accountability under different patterns of AI use?

2 Related Work and Literature Review

2.1 Generative AI and Productivity

Recent field experiments show that generative AI can substantially raise productivity in knowledge work. In customer support and professional writing, AI assistance increases output quality and reduces completion time, with especially large gains for lower-performing workers [1, 2]. These studies treat AI as a production technology and focus on performance *with* AI turned on.

2.2 Skill, Learning, and Practice Design

Classic work on expertise and organizational learning emphasizes that practice structure and feedback shape long-run skill trajectories. Katila's research on search and innovation shows how the timing and diversity of search affect learning and performance [7,8]. Recent work on enabling technologies and ML-based matching points to the importance of how firms structure experimentation and resource allocation when adopting new tools [9, 10]. However, we still lack direct evidence on how daily interactions with AI copilots translate into unaided skill trajectories.

2.3 Algorithmic Management, Authority, and Inequality

Research on algorithms at work documents how algorithmic systems become new loci of control and contestation [3]. Work by Karunakaran and colleagues examines how AI and digital platforms can both reinforce and challenge inequalities in organizations [4]. Valentine and coauthors show how algorithmic coordination and flash organizations restructure collaboration and team formation [5,6]. Together, these streams suggest that AI use profiles could reshape who is perceived as competent, who is trusted with high-stakes work, and how authority is distributed.

2.4 Careers, Institutions, and Opportunity Structures

Eesley's research highlights how talent, experience, and institutional environments jointly shape entrepreneurial and career outcomes [11–13]. These insights transfer naturally to AI-infused organizations, where institutional rules about AI use and evaluation may open or close career paths for different workers. Sako's recent work on generative AI and knowledge work emphasizes the need to integrate technology, tasks, and organizational roles in understanding AI's impact [14].

2.5 Causal Inference in Organizational Field Experiments

Methodologically, the project builds on advances in difference-in-differences and event-study designs for staggered adoption and heterogeneous treatment effects [15–17]. Recent strategy research advocates for careful use of supervised machine learning to improve matching and causal identification in field settings [10].

2.6 Gap and Contribution

Across these literatures, three gaps remain. First, we lack longitudinal evidence on *unaided* skill trajectories under sustained AI assistance; most studies observe performance with AI turned on. Second, existing work rarely links AI use to authority and career outcomes inside organizations, even though algorithmic management research suggests such links are crucial. Third, there is almost no experimental work on *practice protocols*—such as AI-off windows—as design levers that managers can use.

This project directly addresses these gaps by (i) tracking unaided performance over 9–12 months, (ii) connecting AI use profiles to promotion, staffing, and perceived expertise, and (iii) randomizing a temporal practice protocol that organizations could feasibly adopt at scale.

3 Theoretical Framework and Mechanisms

My framework combines ideas from organizational learning, human–AI collaboration, and inequality in the workplace. At a high level, I distinguish between *AI-as-autopilot* and *AI-as-sparring-partner* use profiles, and examine how these interact with practice protocols and organizational context to shape skill trajectories and career outcomes.

First, AI tools change how workers practice and represent knowledge. When AI is used as an autopilot, workers offload cognitively demanding parts of the task (e.g., structuring an argument, writing code, scoping an analysis) and primarily edit or patch AI outputs. This can compress performance differences in the short run but reduce opportunities for effortful retrieval, error-correction, and deep practice, leading to *skill drift* in unaided performance over time. In contrast, AI used as a sparring partner—for critique, explanation, and exploration—can support *skill lift*, especially when workers deliberately generate an initial solution, compare it to AI suggestions, and reflect on discrepancies.

Second, AI reshapes authority and accountability. As AI-augmented outputs become the perceived "default" standard, managers may increasingly evaluate surface fluency and speed rather than underlying expertise, particularly for workers whose skill has quietly drifted. Logs, dashboards, and visible artifacts may privilege those who appear consistently "on top of things" with AI support, shifting who is perceived as an expert, who is trusted with high-stakes assignments, and who becomes a mentor or gatekeeper for others.

Third, these effects are mediated by organizational practices. A Temporal Practice Protocol—scheduled AI-off windows coupled with structured reflection—is a governance lever that intentionally reintroduces unaided practice into AI-intensive workflows. By requiring workers to periodically solve tasks without AI and then reflect on differences between unaided and AI-assisted performance, the protocol can strengthen internal models of the task, maintain calibration about what one knows versus what the AI knows, and make expertise visible again through reflection artifacts (e.g., notes, documentation, internal talks). At the same time, the protocol may impose additional cognitive and time costs, and its benefits may depend on how managers integrate it into evaluation and workload expectations.

Finally, the framework recognizes that AI use is embedded in existing power structures. Differences in role, seniority, team norms, and access to mentorship may create systematic variation in who drifts and who lifts. Without deliberate intervention, AI may thus widen gaps in expertise, visibility, and opportunity, even if short-run productivity gains appear broadly shared. The empir-

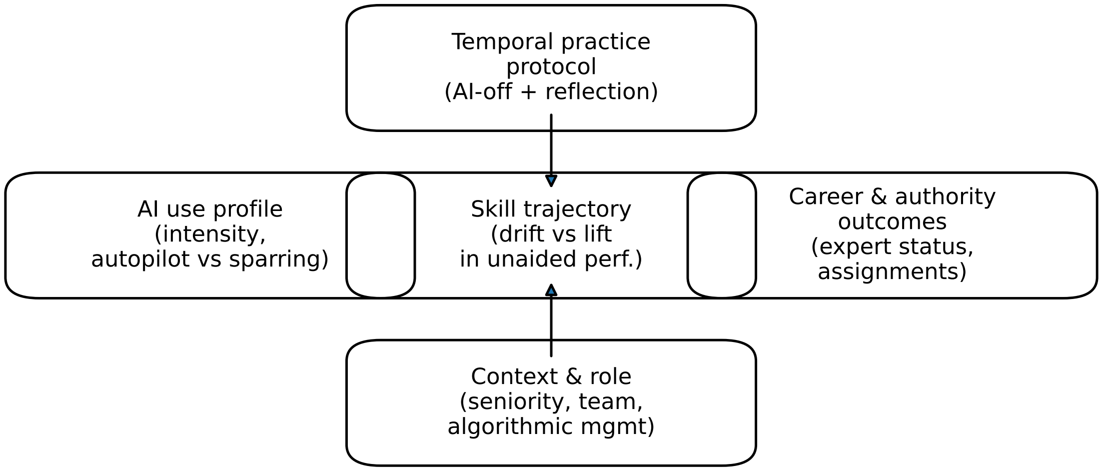

Figure 1: Conceptual model linking AI use profiles, skill trajectories, and career outcomes, with practice protocols and organizational context as key moderators.

ical design treats these organizational and demographic factors as moderators of the relationships between AI use profiles, practice protocols, skill trajectories, and career outcomes.

4 Hypotheses

Operationalizing this framework, I plan to test hypotheses about skill trajectories, visibility, and the role of practice protocols.

4.1 Skill drift and AI use profiles

H1a (Skill drift under autopilot use). Holding baseline skill constant, workers with higher levels of AI-as-autopilot use will show greater declines in unaided task performance over time, relative to workers with lower autopilot use.

H1b (Skill lift under sparring-partner use). Holding baseline skill constant, workers who predominantly use AI as a sparring partner (for critique, explanation, and review) will maintain or improve unaided performance over time, relative to workers who rely more on autopilot use.

4.2 Practice protocols as a moderating mechanism

H2a (Mitigating drift). Among workers with high overall AI use, assignment to the Temporal Practice Protocol will attenuate or reverse skill drift in unaided performance, relative to similar workers in business-as-usual teams.

H2b (Effects on junior workers). The mitigating effect of the Temporal Practice Protocol on skill drift will be strongest for junior workers, who are most likely to rely heavily on AI for core tasks.

4.3 Perceived expertise and career outcomes

H3a (Expert recognition). Controlling for baseline skill and role, workers who preserve or improve their unaided performance will be more likely to be recognized as "go-to" experts by peers and managers.

H3b (High-stakes assignments and promotion). Controlling for baseline skill and role, workers who preserve or improve their unaided performance will be more likely to be staffed on high-stakes projects and to experience more favorable promotion outcomes.

H3c (Mediation by practice). The effect of assignment to the Temporal Practice Protocol on visibility and promotion outcomes will be partially mediated by its effect on unaided performance.

4.4 Psychological experience and ownership

H4 (Ownership and accountability). Assignment to the Temporal Practice Protocol will increase workers' sense of ownership and accountability for their outputs, relative to workers in business-as-usual teams, even when overall AI use remains high.

5 Study Design

I propose a nine- to twelve-month longitudinal field study in AI-augmented work settings (e.g., product, engineering, operations, or customer-support teams using generative AI tools in their daily work). The design has two integrated components: a panel study and a policy intervention.

5.1 Panel tracking of AI use, skill, and visibility

First, I will build a panel of comparable teams that already use AI tools (e.g., code assistants, writing assistants, or internal copilots). For each participant, I will track over time:

- **AI use intensity and mode**: frequency and duration of AI use, types of tasks where AI is used, and whether AI is used as an "autopilot" (first-draft generator) or a "sparring partner" (for review and critique).
- **Unaided task performance**: quarterly "AI-off" benchmark tasks where participants complete realistic work assignments without AI (e.g., drafting a short product spec, writing a piece of code, or diagnosing a case). These are graded with standardized rubrics by independent raters blind to AI use history.
- **Knowledge transfer and mentoring**: contributions to onboarding documents, mentoring interactions, and reusable playbooks (e.g., documentation quality, internal talks, code reviews).
- **Career visibility and outcomes**: assignment to high-stakes projects, promotion decisions, performance ratings, and informal status metrics (e.g., who is consulted as the "go-to" person).
- **Psychological experience**: quarterly surveys on ownership, accountability, confidence, and perceived dependence on AI.

Conceptually, I expect different *AI use profiles*—for example, heavy AI-as-autopilot versus AI-as-sparring-partner—to produce different trajectories of skill drift or skill lift and different patterns of who is seen as an expert.

5.2 Practice protocol intervention: Temporal AI-off windows

Second, I propose a policy intervention: a **Temporal Practice Protocol**. Teams will be randomly assigned to one of two conditions:

1. **Practice condition (AI-off protocol)**: teams adopt scheduled intervals (e.g., one afternoon every two weeks) where AI tools are turned off or strongly discouraged for designated tasks. Participants complete work unaided, then write a short reflection on (a) where they struggled, (b) what they learned about the task structure, and (c) where prior AI outputs helped or hindered their understanding.
2. **Business-as-usual condition**: teams continue with unconstrained AI use, with only measurement (no AI-off requirement).

Within the practice condition, I will also explore small design variations (e.g., individual vs. pair-based reflection; whether reflections are shared with managers or kept private) to understand how these features shape perceived safety, learning, and authority.

Randomizing the introduction of the practice protocol across teams, combined with repeated measurement, will enable causal inference on the effect of the protocol on unaided skill, perceived expertise, and career outcomes.

5.3 Visual overview of the conceptual model and timeline

Figure 1 shows the conceptual model linking AI use profiles, skill trajectories, and career outcomes, with the practice protocol and organizational context as key moderators.

5.4 Power analysis and sample size planning

Because this is a longitudinal, team-level field experiment with repeated measures, power depends on both the number of teams and the number of observations per worker. I plan power calculations around a conservative minimum detectable effect (MDE) on unaided performance of 0.20–0.25 standard deviations, which is smaller than short-run effects in existing GenAI field experiments [1,2].

The basic unit of randomization will be the team. Let J denote the number of teams, with half assigned to the Temporal Practice Protocol and half to business-as-usual AI use. Within each team, I expect $n \approx 6$–10 workers contributing repeated benchmark tasks at baseline $Q0$ and three follow-up quarters $Q1$–$Q3$. As an illustrative target, $J = 40$ teams (20 treatment, 20 control), each with $n = 8$ workers observed over four quarters, yields roughly $N \approx 1{,}280$ worker-quarter observations.[1] Following standard difference-in-differences power approximations,[2] the effective sample size scales with J and the number of time points, but is attenuated by within-worker autocorrelation.

Under plausible assumptions about intra-class correlations (ICC ≈ 0.10 at the team level and 0.50 at the worker level), this design is expected to give at least 80% power to detect a 0.22–0.25 SD reduction in skill drift between treatment and control arms. I will refine these calculations using simulation-based power analysis once I have pilot estimates of variance components from early data, and adjust sample sizes or measurement frequency accordingly.

6 Data, Measures, and Analysis

This section summarizes the main data sources, how key constructs are operationalized, and the analytical strategies used to test the hypotheses. Detailed variable dictionaries and code will be provided in an online appendix.

6.1 Data sources

The study integrates four primary data sources:

1. **AI interaction logs**. Instrumented copilots and collaboration tools generate structured records of prompts, suggestions, accept/override actions, timestamps, and task identifiers. These logs are used to construct AI use features and to infer AI use profiles (autopilot, sparring partner, minimal AI).

2. **Unaided benchmark tasks**. Each quarter, workers complete realistic work-like tasks without AI (e.g., writing a support response, drafting a short product memo, or implementing a coding function). Tasks are administered through a secure platform with AI access disabled for the duration of the task.

3. **Organizational records**. Administrative data include role, tenure, team membership, assignments to projects, performance ratings, promotions, and other career events. Where

[1] $40 \times 8 \times 4 = 1{,}280$ worker-quarter observations.

[2] For example, extensions of two-period DiD formulas to multiple periods as in [15–17].

available, informal visibility indicators (e.g., who is tagged in issue trackers or consulted on design documents) will be derived from collaboration tools.

4. **Surveys and reflections**. Quarterly surveys capture self-reported AI use, ownership, accountability, and perceived dependence on AI, as well as perceptions of fairness and psychological safety. In the Temporal Practice Protocol arm, short reflection prompts during AI-off windows generate additional qualitative text about workers' experiences.

These data sources will be merged using pseudonymous worker and team identifiers into a longitudinal panel at the worker–quarter level.

6.2 Key measures

AI use profiles. Using log-derived features (intensity, acceptance behaviour, initiative, override patterns, and temporal dynamics), I will construct summary indices and cluster workers into AI use profiles (autopilot, sparring partner, minimal AI) as described in Section 6. For regression analyses, both continuous features (e.g., fraction of content authored by AI) and categorical profile indicators will be used.

Unaided performance. Benchmark tasks will be scored with standardized rubrics that assess accuracy, depth, structure, and calibration. For each quarter t, worker i will have an observed score S_{it} and an estimated latent ability θ_{it} from the psychometric models described in the measurement framework (e.g., longitudinal IRT and multilevel growth models).

Knowledge transfer and mentoring. Contributions to onboarding documents, reusable playbooks, and mentoring activities will be measured using a combination of (i) counts of authored or substantially edited documentation, (ii) participation in code reviews and design reviews, and (iii) self- and manager reports of mentoring responsibilities. Where available, text-mining of documentation and review comments will be used to characterize the depth and specificity of contributions.

Visibility and career outcomes. Visibility is operationalized through assignment to high-stakes or central projects (e.g., critical launches, complex incidents), frequency of consultation (mentions or tags in collaboration tools), and peer/manager nominations as "go-to" experts. Career outcomes include promotion events, lateral role changes, and performance-rating bands over the study period. These measures will be constructed from HR systems and collaboration metadata.

Psychological experience. Ownership, accountability, confidence, and perceived dependence on AI will be measured using short scales administered each quarter. Items will be adapted from existing validated measures where possible and refined through cognitive pre-testing. Composite scores will be computed based on reliability analyses, and change scores over time will be used to relate experience to behavioural patterns and treatment status.

Covariates and moderators. Baseline skill, role (e.g., junior vs. senior engineer, product manager, support agent), tenure, prior experience, and demographic characteristics (where available and ethically appropriate) will be included as covariates. These variables will also be used to probe heterogeneity in treatment effects (e.g., differences in skill drift between junior and senior workers).

6.3 Analytical approach

Analyses will proceed in three steps aligned with the hypotheses.

First, to study **skill trajectories** (H1, H2), I will estimate difference-in-differences and event-study models of unaided performance on AI use profiles and protocol assignment, as described using modern difference-in-differences and event-study estimators.

Second, to examine **expert recognition and career outcomes** (H3), I will model visibility and promotion outcomes as functions of protocol assignment, AI use profiles, and changes in unaided performance. For binary or discrete outcomes (e.g., being staffed on a flagship project, promotion to the next level), I will use generalized linear mixed models with worker and team random intercepts. Mediation analyses will assess whether the effect of the practice protocol on visibility and promotion is partially transmitted through its impact on unaided skill.

Third, to understand **psychological experience and mechanisms** (H4 and RQ5), I will relate changes in ownership, accountability, and perceived dependence on AI to treatment status and AI use profiles, and integrate these quantitative patterns with qualitative themes from reflections and interviews. Joint consideration of log data, survey trajectories, and qualitative insights will help distinguish between competing mechanisms (e.g., reduced practice versus increased delegation versus changes in evaluation norms).

Throughout, I will correct for multiple hypothesis testing within outcome families where appropriate, conduct robustness checks such as alternative estimators, pre-trend diagnostics, placebo tests, and leave-one-team-out analyses, and report both point estimates and uncertainty intervals. All analyses will be pre-registered to the extent feasible and implemented using reproducible scripts made available in an online repository.

7 Implementation, Ethics, and Timeline

This section summarizes how the study will be implemented in practice, how data flows from worker–AI interactions into the analysis panel, and how ethical considerations, rival explanations, and timing are addressed.

7.1 Data pipeline and implementation

The empirical design requires an infrastructure that can reliably log AI interactions, link them to benchmark tasks and organizational outcomes, and protect worker privacy. Instrumented AI copilots and collaboration tools will emit structured events (prompts, suggestions, accept/override actions, edit distances, timestamps) through an integration layer that appends pseudonymous worker and team identifiers.

These raw events will be written to an append-only event store and periodically transformed into curated tables in a columnar warehouse (e.g., session-level and worker–quarter feature tables). From there, I will construct:

- worker–period features for AI use profiles (intensity, autopilot vs. sparring, override behaviour),
- benchmark-task tables with scores, rubrics, and rater identifiers, and
- panel datasets linking AI use, unaided performance, visibility, and promotion outcomes.

Figure 2 provides a high-level illustration of this pipeline from worker–AI interaction logs and surveys to the analysis panel used in the main models. Analysis code and ETL scripts will be maintained in a version-controlled repository with a reproducible environment.

7.2 Ethical considerations and IRB strategy

Because the project involves workplace data, ethical design is central. I will work only with organizations willing to adopt a research protocol that includes explicit, informed consent from participants. Monitoring will be limited to work accounts and devices, and data collection will follow

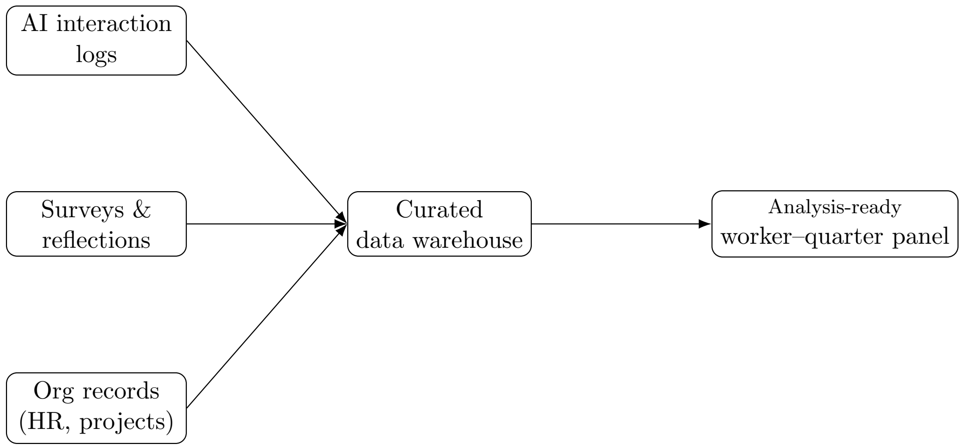

Figure 2: Data pipeline from AI interaction logs, surveys, and organizational records to analysis-ready worker–quarter panel datasets.

principles of data minimization: no keystroke logging, no off-hours tracking, and no collection of content beyond what is necessary to construct AI use features and benchmark outcomes.

To reduce the risk that participation affects employment, I will negotiate data-use agreements specifying that (i) raw research data and derived AI usage features are not shared with line managers or HR for individual evaluation, and (ii) any feedback to organizations is aggregated at team or site level. Personally identifying information will be stored separately under strict access controls, and analytic datasets will use de-identified IDs.

The project will be submitted to the institutional review board with detailed data-flow diagrams, consent forms, and plans for secure storage and eventual data deletion. Where organizations are outside the U.S. or subject to GDPR-like regimes, I will ensure compliance with relevant data-protection regulations in consultation with local counsel.

7.3 Alternative explanations and rival hypotheses

Several rival explanations could account for observed patterns without invoking skill drift or the practice protocol, and the design is intended to address them explicitly.

Selection into AI use. Workers who are already struggling might rely more on AI, creating an association between heavy AI use and lower unaided performance that reflects selection rather than drift. I will address this by controlling for baseline performance and prior evaluations, using worker fixed effects to focus on within-person changes, and comparing trends around the randomized introduction of the practice protocol.

Learning-by-using AI. It is possible that AI enhances understanding by providing high-quality examples or explanations, so that heavy AI use leads to skill lift rather than drift. Hypotheses H1a and H1b already allow for this possibility. I will examine heterogeneity by AI use profile (autopilot vs. sparring partner) and explore mechanisms using reflection logs and interviews.

Managerial favoritism and informal networks. Differences in visibility and promotion could reflect pre-existing managerial preferences or informal networks rather than changes in skill. I will collect data on reporting lines and informal consultation networks (e.g., “who do you go to for help?”) and include these as controls or moderators in the analysis.

Organizational shocks. Organization-wide changes (e.g., restructurings, new tools) might drive outcome trends. Staggered rollout across teams and sites, combined with time fixed effects and robustness checks that drop periods with major shocks, will help separate these from protocol effects.

7.4 Timeline and milestones

Figure 3 summarizes an illustrative three-year timeline for field partnerships, data collection, and analysis. The first year focuses on securing organizational partners, finalizing IRB approval, and piloting benchmark tasks and the practice protocol. The second year emphasizes full rollout of the Temporal Practice Protocol, longitudinal data collection, and interim analyses to monitor data quality. The third year focuses on final analyses, integration of quantitative and qualitative findings, and dissemination to both academic and practitioner audiences.

This staged plan reflects the practical constraints of working with organizations while ensuring sufficient time to observe skill trajectories, authority shifts, and early career outcomes.

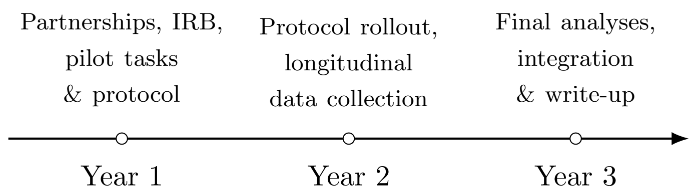

Figure 3: Illustrative three-year timeline for partnerships, data collection, and analysis.

8 Preliminary Evidence and Feasibility

Although this project is designed as a prospective longitudinal field experiment, there is already preliminary evidence that motivates the design and suggests feasibility.

First, I have begun small-scale pilots with convenience samples of knowledge workers (e.g., peers and collaborators) using realistic tasks. In an initial pilot (approximately $N \approx 10$), participants completed two short writing or analysis tasks: one with access to a generative AI assistant and one without. I collected self-reports on how they used AI (autopilot vs. sparring-partner), time spent, and perceived ownership of the final output. Early descriptive patterns were consistent with the proposed framework: participants who described heavy autopilot use reported weaker understanding of the underlying task structure and lower confidence in reproducing the work without AI, whereas those using AI as a sparring partner reported stronger understanding and ownership.

In a second pilot, I prototyped the Temporal Practice Protocol by asking participants to complete a short “AI-off” session followed by a structured reflection (e.g., “What felt harder without AI? What did you notice about your own knowledge? Where did prior AI outputs help or hinder you?”). Participants rated the reflections as valuable and feasible to integrate into normal workflows, suggesting that a scaled-up protocol is practical and acceptable in real teams.

Beyond these small pilots, existing field experiments with generative AI in customer support and professional writing show large short-run gains in output quality and speed, especially for less-experienced workers [1, 2]. These studies demonstrate that organizations are willing to randomize copilot access and instrument fine-grained logs at scale. Research on algorithms at work and AI in organizations documents how algorithmic systems shift control, accountability, and perceptions of expertise [3–6], reinforcing the idea that AI use profiles are likely to have downstream consequences for authority and opportunity, not just productivity.

Finally, through my ongoing industry experience building AI-powered tools for knowledge work, I have already implemented logging pipelines, prompt-governance mechanisms, and A/B tests for AI features in collaboration with engineering teams. This experience provides a practical foundation for the instrumentation and practice protocol proposed here, including familiarity with privacy-preserving logging, integration with existing tools, and coordination with partner organizations.

Together, these pilots, prior studies, and practical experiences provide both empirical and operational evidence that a longitudinal AI-at-work field experiment of the proposed form is feasible, ethically justifiable, and likely to yield informative data on skill trajectories, authority, and career outcomes.

9 Expected Theoretical Contribution (Detailed)

This project is designed to contribute to theory at the intersection of expertise and learning, algorithmic authority, and inequality in AI-intensive organizations. Here I elaborate how the proposed study advances each of these areas.

9.1 Expertise, learning, and practice in AI-augmented work

First, the project speaks to classic theories of expertise and organizational learning by examining how practice structure changes when AI copilots become ubiquitous. Prior work emphasizes effortful retrieval, feedback, and deliberate practice as key engines of skill development. Yet most empirical studies of generative AI treat AI assistance as a one-period productivity shock and measure performance with AI turned on. By contrast, this study focuses on *unaided* performance trajectories over nine to twelve months, distinguishing between AI-as-autopilot and AI-as-sparring-partner use profiles and experimentally introducing a Temporal Practice Protocol that reintroduces AI-off practice windows.

This design allows me to test whether certain patterns of AI use crowd out or complement the kinds of practice that sustain expertise. Evidence that heavy autopilot use leads to measurable skill drift, while sparring-partner use and deliberate practice preserve or enhance unaided performance, would refine theories of learning-by-doing and deliberate practice for an era of AI assistance. Conversely, null or opposite findings would also be informative, indicating where concerns about drift may be overstated and how practice protocols can be simplified.

9.2 Algorithmic authority, visibility, and career trajectories

Second, the project extends work on algorithmic authority and algorithmic management by linking AI use not just to immediate task performance, but to authority, visibility, and career outcomes. Existing research documents how algorithmic systems can become new loci of control and contestation, but typically focuses on settings where algorithms directly allocate work or evaluate performance. Here, AI tools are framed as *assistants* or copilots, yet they still shape what work looks like, what gets logged, and how competence is perceived.

By combining AI-use profiles, benchmark tasks, and visibility measures (e.g., staffing on high-stakes projects, being named as a "go-to" expert), the study examines whether workers whose skills drift under autopilot use nonetheless appear competent in AI-augmented outputs, and whether practice protocols that preserve unaided skill translate into sustained authority and opportunity. Evidence that temporal practice protocols change who is seen as an expert, or that certain AI use profiles systematically shift visibility, would enrich theories of algorithmic authority by showing how authority is co-produced by human practice, AI tools, and organizational evaluation routines.

9.3 Inequality, heterogeneity, and opportunity structures

Third, the project contributes to research on inequality and opportunity structures in organizations by treating AI adoption as a potential amplifier or mitigator of existing gaps. The design explicitly models heterogeneity by role (junior vs. senior), baseline skill, and worker background, and asks whether AI use profiles and practice protocols differentially affect skill trajectories and career outcomes across these groups.

If, for example, junior workers are more likely to adopt autopilot use and experience greater skill drift in the absence of practice protocols, the study would provide a mechanism by which AI tools widen gaps in expertise and promotion despite headline productivity gains. Conversely, if well-designed practice protocols disproportionately benefit junior or historically disadvantaged workers by preserving or accelerating their skill development, this would point to concrete organizational levers for using AI to broaden, rather than narrow, opportunity structures.

9.4 Measurement, governance, and evidence-ready AI systems

Finally, the project advances emerging ideas about *evidence-ready* AI systems by showing how organizational logging, benchmarking, and governance can be designed to support rigorous study and audit of AI's long-run effects. The proposed instrumentation—structured AI interaction logs, periodic unaided benchmark tasks, and traceable links to staffing and promotion data—serves a dual purpose: enabling causal inference about the effects of practice protocols, and illustrating how organizations can build AI infrastructures that make questions about skill, authority, and inequality empirically tractable.

By demonstrating a concrete architecture for logging, measurement, and governance in AI-augmented teams, the study offers a template for future work that treats AI copilots not just as black-box productivity tools, but as components of sociotechnical systems whose effects on capability and opportunity can be measured and governed. This, in turn, contributes to broader conversations in management science and engineering about how to design AI-intensive organizations that are both productive and just.

10 Conclusion

This project asks whether AI copilots produce skill drift or skill lift, how practice protocols can shape these trajectories, and who gains or loses authority and opportunity in AI-intensive organizations. While recent field experiments document sizable short-run productivity gains from generative AI, we know far less about how sustained AI assistance affects unaided capabilities, expertise recognition, and career paths over time.

I propose a longitudinal field study that combines instrumented AI-use logs, quarterly unaided benchmark tasks, and organizational data on visibility and promotion, together with a randomized Temporal Practice Protocol that reintroduces AI-off practice windows and structured reflection. This design makes it possible to distinguish between AI-as-autopilot and AI-as-sparring-partner use profiles, estimate the causal effects of practice protocols on skill trajectories, and examine how these dynamics play out across roles and worker backgrounds.

Theoretically, the study aims to refine accounts of expertise and learning in AI-augmented work, extend theories of algorithmic authority to settings where AI is framed as an assistant rather than an evaluator, and illuminate mechanisms through which AI adoption may widen or narrow existing inequalities. Methodologically, it demonstrates how organizations can build evidence-ready AI infrastructures that support rigorous evaluation of AI's long-run effects.

By treating AI copilots, organizational practices, and measurement infrastructures as jointly designable elements of sociotechnical systems, the project aspires to generate evidence that is both analytically rigorous and practically useful for organizations seeking to harness AI while preserving expertise, accountability, and fair opportunity.

References

[1] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, vol. 381, no. 6654, pp. 187–192, 2023.

[2] E. Brynjolfsson, D. Li, and L. R. Raymond, "Generative AI at work," *Quarterly Journal of Economics*, vol. 140, no. 2, pp. 889–948, 2025.

[3] K. C. Kellogg, M. A. Valentine, and A. Christin, "Algorithms at work: The new contested terrain of control," *Academy of Management Annals*, vol. 14, no. 1, pp. 366–410, 2020.

[4] A. Karunakaran, H. Rahman, and coauthors, "Artificial intelligence at work: An integrative perspective on the impact of AI on workplace inequality," *Academy of Management Annals*, 2025, forthcoming.

[5] M. A. Valentine, D. Retelny, A. To, N. Rahmati, T. Doshi, and M. S. Bernstein, "Flash organizations: Crowdsourcing complex work by structuring crowds as organizations," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 3523–3537.

[6] M. A. Valentine and coauthors, "How algorithms can conflict with organizational structures," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, no. CSCW, 2024, article no. to be assigned.

[7] R. Katila, "New product search over time: Past ideas in their prime?" *Academy of Management Journal*, vol. 45, no. 5, pp. 995–1010, 2002.

[8] R. Katila and E. L. Chen, "Effects of search timing on innovation: The value of not being in sync with rivals," *Administrative Science Quarterly*, vol. 53, no. 4, pp. 593–625, 2008.

[9] J. M. Rathje and R. Katila, "Enabling technologies and the role of private firms: A machine learning matching analysis," *Strategy Science*, vol. 6, no. 1, pp. 5–21, 2021.

[10] J. M. Rathje, R. Katila, and P. Reineke, "Making the most of AI and machine learning in organizations and strategy research: Supervised machine learning for matching," *Strategic Management Journal*, 2024, forthcoming.

[11] C. E. Eesley and E. B. Roberts, "Are you experienced or are you talented? when does innate talent versus experience explain entrepreneurial performance?" *Strategic Entrepreneurship Journal*, vol. 6, no. 3, pp. 207–219, 2012.

[12] C. Eesley, "Institutional barriers to growth: Entrepreneurship, human capital and institutional change," *Organization Science*, vol. 27, no. 5, pp. 1290–1306, 2016.

[13] C. E. Eesley, R. N. Eberhart, B. R. Skousen, and J. L. C. Cheng, "Institutions and entrepreneurial activity: The interactive influence of misaligned formal and informal institutions," *Strategy Science*, vol. 3, no. 2, pp. 393–407, 2018.

[14] M. Sako, "How generative AI fits into knowledge work," *Communications of the ACM*, vol. 67, no. 5, pp. 18–21, 2024.

[15] B. Callaway and P. H. C. Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, vol. 225, no. 2, pp. 200–230, 2021.

[16] L. Sun and S. Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, vol. 225, no. 2, pp. 175–199, 2021.

[17] K. Borusyak, X. Jaravel, and J. Spiess, "Revisiting event-study designs: Robust and efficient estimation," *Review of Economic Studies*, vol. 91, no. 6, pp. 3253–3287, 2024.