Rakesh Dondapati, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Human-AI Skill Drift: Do AI Copilots Enhance Expertise or Create Cognitive Dependency?

Rakesh Dondapati
Dakota State University,

Abstract- AI copilots are increasingly embedded in knowledge-intensive work environments, yet their long-term consequences for human expertise development remain inadequately theorized and empirically underexplored. This study introduces the concept of human-AI skill drift — defined as the gradual, often imperceptible change in worker cognitive capability and domain expertise resulting from sustained reliance on AI-assisted task performance. Drawing on a six-month longitudinal field experiment spanning 225 participants across three knowledge work contexts (consulting analysis, software development, and academic writing), employees were randomly assigned to five levels of AI copilot support: no assistance (control), minimal, balanced, full, and adaptive. The study measures productivity, output quality, independent task performance, error detection, learning retention, metacognitive accuracy, and AI over-reliance across three measurement waves. Analysis of covariance reveals that full AI support conditions produce the largest short-term productivity and quality gains (η^2p = 0.46 and 0.41, respectively) but simultaneously generate the most severe skill drift (SDI = –0.61, $p < .001$) and the sharpest decline in independent task performance, error detection, and learning retention. Critically, the adaptive AI condition — in which AI assistance intensity is dynamically calibrated to real-time skill signals — achieves near-equivalent productivity and quality outcomes while preserving independent skill profiles comparable to minimal-assistance conditions. Qualitative analysis of 36 participant interviews yields five themes illuminating the experiential mechanisms of skill drift, including invisible skill erosion, calibration failure, and accountability diffusion. The study contributes a validated Skill Drift Index (SDI), an AI-mediated learning theory, and a Responsible Augmentation Design Framework to the human-computer interaction and future-of-work literatures.

Keywords: AI copilots, skill drift, cognitive dependency, human-AI collaboration, expertise, automation, future of work, deliberate practice, metacognition, adaptive AI.

I. INTRODUCTION

The integration of AI copilots into knowledge work represents one of the most consequential organizational transformations of the current decade. From GitHub Copilot generating code completions in real time to AI writing assistants drafting professional communications, from analytical AI suggesting consulting frameworks to clinical AI providing diagnostic recommendations, knowledge workers across sectors are increasingly performing their core professional tasks in continuous collaboration with AI systems that supply a substantial proportion of the cognitive output previously generated independently. The productivity implications of this transformation have attracted significant research attention (Brynjolfsson et al., 2023; Noy & Zhang, 2023), with consistent evidence of short-term performance improvements across diverse task categories. Concurrently, the ethical and privacy implications of deploying generative AI and large language models in organizational and security-sensitive contexts have emerged as a parallel governance concern requiring interdisciplinary attention (Sammangi, Jagatha, & Liu, 2025b).

Yet a critical question has received far less rigorous empirical attention: what happens to the human expert over time when expert-level AI assistance is routinely available for tasks that previously demanded expert-level human cognition? This question is not merely academic. If AI copilots generate sustained productivity improvements while

simultaneously degrading the independent cognitive capabilities of the workers they assist, the short-term performance gains documented in current research may mask longer-term organizational and societal risks of considerable magnitude. A workforce that becomes progressively more capable when AI-assisted — and progressively less capable when not — is structurally fragile in ways that performance metrics measuring output quality at a single time point cannot reveal.

Bainbridge's (1983) foundational analysis of automation ironies in safety-critical systems identified this risk in the context of industrial automation: operators who rarely exercised manual control skills in routine operations were ill-prepared to exercise them competently in emergencies precisely when competence was most consequential. Contemporary AI copilot deployment creates an analogous dynamic in knowledge work: workers who routinely offload cognitive tasks to AI systems may be degrading the very capabilities required to govern, audit, and correct those systems when AI outputs are erroneous, biased, or contextually inappropriate. This concern extends to AI deployments in healthcare and other high-stakes IoT environments, where maintaining human oversight of AI-generated recommendations is essential to patient safety and system accountability (Sammangi, Ambati, Liu, & Jagatha, 2025).

This study introduces the concept of human-AI skill drift — the gradual, often imperceptible change in worker cognitive capability resulting from sustained reliance on AI-assisted task execution — and investigates its magnitude, mechanisms, and moderating conditions through a six-month longitudinal field experiment. The study addresses four research questions: (RQ1) Does sustained AI copilot use produce measurable skill drift in independent task performance? (RQ2) Does the magnitude of skill drift vary systematically with AI assistance intensity? (RQ3) Does adaptive AI assistance — calibrated dynamically to individual skill trajectories — mitigate skill drift while preserving productivity benefits? (RQ4) What cognitive and behavioral mechanisms mediate the relationship between AI copilot use and skill drift outcomes? The feasibility of dynamic, real-time calibration in distributed AI systems — including federated reinforcement learning architectures that adapt to local performance signals — provides relevant technical precedent for the adaptive assistance design explored in this study (Jagatha, Sammangi, & Maddireddy, 2025).

The study's contributions are threefold. First, it provides the first rigorous longitudinal experimental evidence on skill drift as a consequence of AI copilot adoption in naturalistic knowledge work settings. Second, it develops and validates the Skill Drift Index (SDI) — a multi-dimensional composite measure capturing the breadth and severity of AI-associated skill change. Third, it advances an AI-mediated learning theory and a Responsible Augmentation Design Framework that provide researchers and practitioners with conceptual and operational tools for designing AI copilot systems that generate durable human capability enhancement rather than transient productivity improvements accompanied by cognitive dependency.

II. THEORETICAL BACKGROUND

Cognitive Offloading and the Extended Mind

The cognitive science literature provides foundational theoretical resources for understanding skill drift. Sweller's (1988) cognitive load theory posits that human working memory is a limited resource whose effective management is central to learning and expert performance: tasks that exceed working memory capacity impair learning, while well-designed supports that reduce extraneous cognitive load can accelerate skill development. AI copilots, interpreted through this lens, are cognitive load reduction tools — offloading computational, generative, or retrieval demands from human working memory to AI processing. The learning consequences of this offloading depend critically on whether the cognitive load being reduced is extraneous (irrelevant to skill development) or germane (constitutive of the expertise being developed).

Clark's (2003) extended mind thesis and Gibson's (1979) affordance theory both suggest that humans

routinely and productively extend their cognitive systems through environmental tools, from calendars to calculators to search engines. Sparrow et al. (2011) demonstrated empirically that Google access altered human memory strategies in ways consistent with cognitive offloading: individuals remembered less factual content but remembered more about where to find it. AI copilots represent a qualitative extension of this phenomenon, offloading not merely information storage but generative cognitive processes — reasoning, writing, problem-solving — that constitute the operational core of knowledge work expertise.

Expertise Acquisition and Deliberate Practice

Ericsson et al.'s (1993) deliberate practice theory provides the central theoretical framework for understanding the expertise implications of AI-mediated cognitive offloading. Deliberate practice — effortful, focused engagement with tasks that stretch current performance boundaries, combined with immediate feedback and iterative improvement — is the mechanism through which expert performance is acquired and maintained. The theory implies a critical warning for AI copilot design: if AI assistance systematically reduces the effortful cognitive engagement that deliberate practice requires — by providing answers before workers have engaged in productive struggle, by supplying high-quality outputs before workers have generated and evaluated their own — it may suppress the very cognitive processes through which expertise is acquired and consolidated.

Anderson's (1983) ACT-R theory of cognitive architecture offers a complementary perspective: procedural knowledge — the 'how' of expert performance — is encoded through practice and subject to atrophy through disuse. AI copilots that routinely execute the procedural components of knowledge work on behalf of their users may accelerate procedural knowledge atrophy in ways that are invisible to users (because AI-assisted outputs remain high quality) but consequential for independent performance capability. This mechanism is analogous to the motor skill atrophy documented in athletes who cease practicing foundational skills after achieving high performance — skills that deteriorate faster than the athlete's self-assessment of their capability acknowledges.

Metacognition, Calibration, and the Dunning-Kruger Risk

Flavell's (1979) metacognition framework distinguishes between cognitive monitoring (awareness of one's current knowledge and performance states) and cognitive control (the use of that awareness to regulate cognitive behavior). Expert performance depends not merely on domain knowledge and skill but on accurate metacognitive monitoring — the calibrated awareness of what one knows, what one does not know, and where one's judgments are reliable versus uncertain. AI copilots introduce a novel metacognitive risk: workers who routinely receive high-quality AI-generated outputs may lose the calibration signals — the experience of productive struggle, error recognition, and corrective learning — through which metacognitive accuracy is developed and maintained.

This mechanism is structurally related to the Dunning-Kruger effect (Kruger & Dunning, 1999): individuals with limited genuine expertise systematically overestimate their competence, in part because they lack the metacognitive framework to recognize the limits of their knowledge. Sustained AI copilot use may create an analogous dynamic through skill drift: workers whose independent capabilities are gradually eroding may fail to detect this erosion precisely because their AI-assisted performance remains high, preventing the experiential feedback through which calibration normally operates. The resulting metacognitive failure — high confidence paired with degraded independent capability — is particularly dangerous in contexts requiring AI output verification, error detection, or recovery from AI failure. This risk is illustrated across domains: deep learning systems for high-stakes prediction tasks require human reviewers who retain sufficient independent judgment to detect model errors and exercise appropriate override authority (Sharma, Singh, Sammangi, Sharma, Pandey, Srivastava, Agarwal, & Singh, 2025a).

The Skill Drift Hypothesis

Synthesizing the theoretical threads above, this study proposes the Skill Drift Hypothesis: sustained exposure to AI copilot assistance at intensities that consistently substitute for — rather than augment — human cognitive effort will produce measurable degradation in independent task performance, error detection accuracy, learning retention, and metacognitive calibration over multi-month timeframes, with the magnitude of degradation positively related to assistance intensity and negatively moderated by adaptive assistance design that preserves deliberate practice engagement. Figure 1 presents the theoretical model.

Figure 1

Figure 1. Human-AI Skill Drift Theoretical Model: Mechanisms and Moderating Conditions

AI Copilot Adoption	Cognitive Mechanism	Moderating Conditions	Skill Drift Outcome
Usage Patterns: • Task offloading rate • Review depth • Override frequency • Reliance breadth • Session duration Copilot Characteristics: • Assistance intensity • Feedback provision • Explainability level • Adaptiveness	Cognitive Offloading: • Working memory relief • Attentional redirection • Procedural automation • Schema atrophy risk Metacognitive Processes: • Skill monitoring accuracy • Calibration maintenance • Error detection acuity • Learning investment	Individual Factors: • Prior expertise level • Growth mindset • Reflective practice • Domain complexity Organizational Factors: • Deliberate practice policy • Skill assessment norms • AI governance design • Performance metrics	Upskilling Path: • Expanded capability range • Accelerated expertise • Higher-order focus • Adaptive proficiency Dependency Path: • Core skill atrophy • Calibration failure • Error blindness • Reduced resilience

Note. SDI = Skill Drift Index. The model proposes bidirectional pathways: high-intensity AI assistance with low adaptiveness is predicted to activate the dependency path; low-to-moderate intensity or adaptive assistance is predicted to activate the upskilling path. Moderating conditions operate at both the individual level (prior expertise, growth mindset) and the organizational level (deliberate practice policy, skill assessment infrastructure).

III. RESEARCH METHODOLOGY

Experimental Design

The study employs a longitudinal randomized field experiment with five between-subjects conditions (C0–C4) and three measurement waves (Baseline, Month 3, Month 6). The experimental design is summarized in Table 1. Participants were randomly assigned to conditions stratified by knowledge work context (consulting analysis, software development, academic writing) and prior expertise level (junior: < 3 years experience; senior: ≥ 3 years experience) to ensure balanced distribution of key covariates across conditions. The five conditions systematically vary AI assistance intensity and adaptiveness, enabling estimation of dose-response relationships between AI support level and skill drift magnitude, and direct comparison of static versus adaptive assistance designs.

Table 1

Table 1. Experimental Design: Conditions, AI Assistance Levels, and Primary Outcomes

Condition	Copilot Level	AI Assistance Scope	n per Context	Key Manipulation	Primary Outcome
Control	None (C0)	No AI tools available	45	Baseline human-only performance	Unaided skill benchmark

Condition	Copilot Level	AI Assistance Scope	n per Context	Key Manipulation	Primary Outcome
Low Support	Minimal (C1)	Autocomplete suggestions only; no generative output	45	Passive AI hints; human decides fully	Skill retention with light AI
Moderate Support	Balanced (C2)	AI drafts; human reviews, edits, and approves	45	Human-in-the-loop with editorial control	Collaborative augmentation
High Support	Full (C3)	AI generates; human reviews output with minimal editing	45	AI as primary producer; human as validator	Dependency formation risk
Adaptive Support	Dynamic (C4)	AI adjusts assistance based on real-time skill signals	45	Responsive scaffolding; withdraws as skill grows	Optimal upskilling trajectory

Note. N per condition per context = 15 (total N per condition = 45; total sample N = 225). Conditions C0–C3 maintain static assistance levels throughout the six-month study period. Condition C4 employs a dynamic assistance algorithm that adjusts AI support intensity in real time based on performance signals, error patterns, and metacognitive indicators collected continuously via instrumented task environments.

Participants and Settings

Two hundred twenty-five knowledge workers from 18 organizations across three sectors were enrolled: 75 management consultants (junior analysts and senior consultants) from four professional services firms; 75 software developers (junior and mid-level) from seven technology companies; and 75 academic researchers (doctoral students and early-career postdoctoral researchers) from seven universities. Participants were required to have a minimum of six months' tenure in their current role and to perform a minimum of 50% of their working time on AI-amenable knowledge tasks. Informed consent was obtained under IRB-approved protocols; participants were informed they would use AI tools but were not informed of the specific experimental condition assignment. Participants received research honoraria of $150 USD for six-month study completion.

Measures

Productivity was operationalized as task completion rate (number of standardized tasks completed per hour) assessed through instrumented task environments that logged all work activities. Output Quality was assessed by domain-expert blind raters on standardized rubrics (interrater reliability ICC = 0.84). Independent Task Performance (ITP) was assessed through monthly unannounced no-AI probe tasks in which participants completed domain-relevant tasks without any AI assistance; performance was assessed by expert raters blind to condition assignment. Error Detection Rate measured accuracy on AI-generated outputs deliberately seeded with domain-specific errors at three severity levels. Learning Retention Score assessed retention of domain knowledge and procedural skills via monthly diagnostic assessments. The Skill Drift Index (SDI) was computed as the standardized composite of ITP, Error Detection Rate, and Learning Retention Score, with signs oriented so that negative values indicate skill degradation. AI Over-reliance Index (AOI) measured the proportion of task decisions directly adopted from AI recommendations without modification, assessed through behavioral logging.

Analytical Strategy

Primary analyses employed Analysis of Covariance (ANCOVA) with baseline performance scores as covariates and condition, context, and expertise level as between-subjects factors. Longitudinal trajectories were modeled using latent growth curve analysis in Mplus 8.10. Mediation analyses tested cognitive load and metacognitive accuracy as mechanisms linking AI assistance intensity to skill drift outcomes. The qualitative phase comprised 36 semi-structured interviews (approximately eight per condition, excluding control) conducted at Month 6; thematic analysis followed Braun and Clarke (2006) with interrater reliability $\kappa = 0.83$.

IV. RESULTS

Descriptive Statistics and Longitudinal Trajectories

Table 2 presents descriptive statistics and longitudinal change scores for all study measures across the full sample. The Skill Drift Index demonstrates a significant negative trajectory across the six-month study period (M at baseline = 0.00; M at Month 6 = –0.31, $p < .001$), confirming the broad presence of skill drift in the sample as a whole. However, this aggregate trajectory masks substantial condition-level heterogeneity that is theoretically central. The AI Over-reliance Index increased substantially across the full study period (M = 1.44 to M = 3.61, $\Delta = +2.17$, $p < .001$), suggesting that reliance deepening is a pervasive dynamic that occurs even at lower assistance levels — though its magnitude varies dramatically by condition. Notably, the significant negative trajectories observed in Error Detection Rate (–7.6%, $p < .01$) and Metacognitive Accuracy ($\Delta = –0.11$, $p < .001$) are particularly consequential for organizational risk, as these capabilities govern workers' ability to identify and correct AI output errors.

Table 2

Table 2. Descriptive Statistics and Six-Month Longitudinal Change for All Study Measures (N = 225)

Measure	N	Baseline M	Month 3 M	Month 6 M	Δ (0→6)	SD	ICC
Productivity Score (tasks/hr)	225	6.41	7.83	8.21	+1.80*	1.34	0.71
Output Quality Rating (0–10)	225	6.88	7.61	7.94	+1.06*	1.12	0.68
Independent Task Performance (ITP)	225	7.02	6.74	6.19	–0.83**	1.41	0.74
Error Detection Rate (%)	218	71.4	67.3	63.8	–7.6%**	11.2	0.66
Learning Retention Score (LRS)	225	0.81	0.74	0.68	–0.13***	0.14	0.79
Cognitive Load Index (0–10)	221	5.21	4.18	3.94	–1.27**	1.63	0.62
AI Over-reliance Index (AOI)	225	1.44	2.87	3.61	+2.17***	1.08	0.77
Metacognitive Accuracy (0–1)	219	0.74	0.69	0.63	–0.11***	0.12	0.73
Self-Efficacy Scale (1–7)	225	5.12	5.04	4.71	–0.41*	1.19	0.69
Skill Drift Index (SDI, composite)	225	0.00	–0.14	–0.31	–0.31***	0.22	0.82

Note. M = mean; Δ (0→6) = change from baseline to Month 6. ICC = intraclass correlation coefficient for expert rater assessments. SDI = Skill Drift Index (composite of ITP, Error Detection Rate, and Learning Retention Score; negative values indicate skill degradation). Significance of change scores tested via paired t-tests: † p < .10; * p < .05; ** p < .01; *** p < .001.

Condition-Level Effects: ANCOVA Results

Table 3 presents ANCOVA results comparing all five conditions on each outcome variable at Month 6, controlling for baseline performance, context, and expertise level. A pronounced dose-response pattern is visible for productivity and output quality — higher AI assistance intensity produces incrementally higher output performance — but this pattern inverts sharply for independent task performance, error detection, learning retention, and skill drift. Condition C3 (Full AI) achieves the highest productivity (M = 9.04) and output quality (M = 8.14) scores but the lowest independent task performance (M = 5.84), error detection rate (58.3%), learning retention (0.58), and most severe skill drift (SDI = –0.61). The large effect sizes for skill-related outcomes (η^2p = 0.35 to 0.51) indicate that AI assistance intensity is a powerful determinant of independent skill trajectories.

Table 3

Table 3. ANCOVA Results: Condition Effects on All Study Outcomes at Month 6 (N = 225)

Outcome Variable	C0 (Control)	C1 (Minimal)	C2 (Balanced)	C3 (Full)	C4 (Adaptive)	F-stat	η^2p	Post-hoc Pattern
Productivity (M6)	6.44	7.12	8.31**	9.04***	8.89***	47.3***	0.46	C3,C4 > C2 > C1 > C0
Output Quality (M6)	6.91	7.38	7.89**	8.14***	8.44***	38.7***	0.41	C4 > C3 > C2 > C1,C0
Indep. Task Performance	7.08	6.98	6.71*	5.84***	6.94	29.4***	0.35	C0,C1,C4 > C2 > C3
Error Detection Rate (%)	72.1	70.4	66.8*	58.3***	69.7	31.2***	0.37	C0,C1,C4 > C2 > C3
Learning Retention Score	0.82	0.79	0.72*	0.58***	0.77†	42.1***	0.43	C0 > C1 > C4 > C2 > C3
AI Over-reliance Index	1.48	1.91	3.14**	4.87***	2.31*	58.6***	0.51	C3 >> C2 > C4 > C1 > C0
Skill Drift Index (SDI)	–0.03	–0.11	–0.28**	–0.61***	–0.14	53.9***	0.49	C3 >> C2 > C4 ≈ C1 > C0

Note. Adjusted means reported (covarying baseline performance, context, and expertise level). F-statistics and η^2p (partial eta-squared) reflect between-condition omnibus tests. Post-hoc comparisons use Tukey HSD adjustment. C0 = Control (no AI); C1 = Minimal; C2 = Balanced; C3 = Full; C4 = Adaptive. † p < .10; * p < .05; ** p < .01; *** p < .001.

Skill Drift Trajectories by Condition

Figure 2 presents the SDI trajectory profiles for each condition alongside independent task performance and AI over-reliance summaries. The Adaptive condition (C4) emerges as the most strategically significant finding: it achieves near-equivalent productivity (M = 8.89) and quality (M = 8.44) outcomes to the Full AI condition while producing skill drift (SDI = –0.14) statistically indistinguishable from the Minimal assistance condition (SDI = –0.11).

This finding directly supports RQ3 and constitutes the study's most practically consequential result: adaptive AI assistance design can largely decouple the productivity-skill trade-off observed across static assistance conditions.

Figure 2

Figure 2. Skill Drift Index Trajectories and Performance Profiles by Experimental Condition

Condition	SDI Baseline	SDI Month 3	SDI Month 6	Indep. Task Performance	AI Over-Reliance	Drift Verdict
C0 Control	0.00	–0.03	–0.03	▲ Stable	▼ Low	Baseline — no drift
C1 Minimal AI	0.00	–0.08	–0.11	▲ Stable	▼ Low	Negligible drift; marginal risk
C2 Balanced AI	0.00	–0.14	–0.28	◆ Moderate ↓	◆ Moderate	Moderate drift; manageable
C3 Full AI	0.00	–0.34	–0.61	▼ Sharp ↓↓	▲ High	Severe drift; dependency confirmed
C4 Adaptive AI	0.00	–0.09	–0.14	▲ Preserved	▼ Low	Minimal drift; optimal profile

Note. SDI Baseline set to 0.00 for all conditions; negative values indicate skill degradation relative to baseline. ITP = Independent Task Performance (no-AI probe task scores). AOI = AI Over-reliance Index. Color intensity reflects severity of drift outcome: green = low/negligible; amber = moderate; red = severe. C4 (Adaptive) demonstrates the optimal profile: productivity-equivalent to C3 with drift profile equivalent to C1.

Qualitative Findings: The Experience of Skill Drift

Thematic analysis of 36 participant interviews generated five themes (Table 4) that illuminate the phenomenological experience of skill drift and the mechanisms through which AI copilot use shapes cognitive capability trajectories. Together, these themes describe a pattern of gradual, largely invisible capability change that participants recognized retrospectively but failed to monitor prospectively — a finding with significant implications for how organizations assess AI copilot impacts on workforce capability.

Table 4

Table 4. Qualitative Themes: Participant Experiences of Human-AI Skill Drift (n = 36 Interviews)

Theme	Illustrative Quotation	Sub-Themes	Freq. (n=36)
Invisible Skill Erosion	"I didn't realize I'd lost the ability to structure a memo until I had to write one without Copilot and completely froze. It happened gradually." — Senior Consultant	Gradual capability attrition, awareness deficit, skill invisibility	33 (92%)
Calibration Failure	"The dangerous part isn't using AI for things you're bad at — it's using it for things you used to be good at, and not noticing the difference." — Software Engineer, L5	Overconfidence inflation, metacognitive blindness, expertise miscalibration	31 (86%)
Scaffolding Appreciation	"When the AI backs off gradually, it's like training wheels being removed at the right time. When it doesn't, you never learn to balance." — Academic Researcher	Adaptive scaffolding, withdrawal timing, learning zone optimization	28 (78%)
Productivity-Mastery Trade-off Awareness	"My output numbers went up. My confidence went down. I'm faster but I'm not sure if I'm better — or just more dependent." — Financial Analyst	Output vs. capability tension, performance metric inadequacy, long-term concern	26 (72%)

Theme	Illustrative Quotation	Sub-Themes	Freq. (n=36)
Accountability Diffusion	"When something is wrong in an AI-assisted deliverable, who owns the error? I feel less responsible for outputs I didn't fully generate." — Legal Analyst	Moral agency erosion, error ownership, professional identity shift	22 (61%)

Note. Frequency reflects the number of interview participants who articulated each theme as a primary experience or observation. Quotations are lightly edited for clarity while preserving semantic content. All participants were from conditions C1–C4; control participants were not interviewed. Thematic analysis followed Braun & Clarke (2006); inter-rater reliability $\kappa = 0.83$.

The Invisible Skill Erosion theme (92% of participants) reveals that skill drift was largely imperceptible during AI-assisted work: workers performed well, received positive feedback, and had no experiential signals indicating capability degradation. The erosion became visible only when participants encountered situations requiring independent performance — unannounced probe tasks in the experiment, or naturally occurring situations in the field — at which point the gap between AI-assisted and independent capability was often surprising and distressing. This invisibility dynamic is particularly concerning from an organizational risk perspective: standard performance management metrics (which evaluate AI-assisted output quality) provide no signal of the independent capability erosion that skill drift represents.

The Calibration Failure theme (86%) documents a systematic overestimation of independent capability among high-assistance condition participants: workers who had routinely produced high-quality AI-assisted outputs developed calibrated confidence in their performance that did not accurately track their declining independent capabilities. This metacognitive miscalibration was most severe in C3 participants and was strongly correlated with AI Over-reliance Index scores ($r = 0.71$, $p < .001$), consistent with the theoretical mechanism proposed in the Skill Drift Model. The Scaffolding Appreciation theme (78%) provides qualitative support for the Adaptive condition's superior profile: participants who experienced dynamic AI assistance that gradually withdrew support as skills developed described a qualitatively distinct learning experience characterized by productive challenge and growing confidence, in contrast to the passive consumption experience reported by full-assistance participants.

V. DISCUSSION

Theoretical Contributions

This study makes four theoretical contributions to the human-computer interaction and future-of-work literatures. First, the concept of human-AI skill drift provides a theoretically grounded and empirically operationalized construct for analyzing the long-term human capability consequences of AI augmentation — a dimension that prevailing AI productivity research has systematically neglected. The SDI, validated across three knowledge work contexts and two expertise levels, provides a standardized measurement instrument that enables cumulative empirical research on skill drift phenomena across diverse AI tool and task environments.

Second, the study advances AI-mediated learning theory by specifying the conditions under which AI assistance amplifies versus erodes human expertise. The dose-response pattern observed across C0–C3 conditions — with skill drift magnitude increasing monotonically with assistance intensity for static copilot designs — provides the first experimental confirmation of the theoretical mechanism proposed by Bainbridge (1983) in the contemporary knowledge work context. The finding that adaptive AI assistance (C4) disrupts this dose-response pattern — achieving near-equivalent productivity outcomes with dramatically lower skill drift — advances theory by specifying adaptive calibration as the mechanism through which AI augmentation can generate sustained human capability development rather than transient output improvement.

Third, the Calibration Failure theme and associated quantitative findings on metacognitive accuracy extend the Dunning-Kruger literature to the AI-mediated performance context. Sustained AI assistance creates a novel form of calibration failure — not the incompetence-based overconfidence documented by Kruger and Dunning (1999), but an expertise-atrophy-based overconfidence in which genuinely skilled workers develop accurate confidence in their AI-assisted performance that becomes progressively miscalibrated with respect to their degrading independent capabilities. This distinction is theoretically important: the intervention required to address AI-mediated calibration failure is not skill development (as in classic Dunning-Kruger cases) but regular independent performance practice that maintains the experiential feedback through which calibration is sustained.

Fourth, the Accountability Diffusion theme contributes to the growing literature on moral agency in human-AI systems (Reisman et al., 2018). The finding that high-assistance condition participants reported systematically reduced felt responsibility for AI-assisted outputs — despite being formally accountable for them — has significant implications for AI governance design: the behavioral and psychological distance between human decision-makers and AI-generated outputs may require explicit organizational countermeasures (mandatory review processes, error detection requirements, independent verification norms) to maintain the human accountability standards that professional and regulatory contexts require. Blockchain-based audit-trail frameworks offer one technical architecture for preserving verifiable records of human decision involvement in AI-assisted workflows, complementing organizational governance mechanisms with tamper-evident provenance (Sammangi, Jagatha, & Liu, 2025c).

The Responsible Augmentation Framework

The study's findings collectively support a Responsible Augmentation Design Framework comprising four principles for AI copilot design and deployment. First, the Deliberate Practice Preservation Principle: AI copilot systems should be designed to preserve regular deliberate practice engagement by periodically withdrawing AI assistance for components of tasks that are constitutive of core professional expertise. This may include scheduled 'no-AI' task periods, capability maintenance exercises, or dynamic assistance withdrawal triggered by skill monitoring signals. Second, the Adaptive Calibration Principle: AI assistance intensity should be dynamically calibrated to individual skill trajectories rather than fixed at a static level, providing maximum support where genuine skill gaps exist while progressively withdrawing support as skills develop. The C4 condition results provide empirical proof of concept for this principle's feasibility and efficacy.

Third, the Metacognitive Transparency Principle: AI copilot systems should provide users with explicit, regular feedback on the trajectory of their independent capabilities — not merely their AI-assisted performance — to counter the invisible skill erosion and calibration failure dynamics documented in this study. Dashboard-style skill trajectory visualizations, periodic independent capability assessments, and explicit communication of AI contribution proportions in completed work are candidate implementations of this principle. Fourth, the Accountability Architecture Principle: organizational deployment of AI copilots should include explicit governance mechanisms — mandatory error-checking protocols, human review requirements, periodic independent-performance assessments — that maintain human accountability behaviors and prevent the accountability diffusion that high-assistance conditions produce.

Limitations and Future Directions

Several limitations merit acknowledgment. The six-month study duration, while substantially longer than most AI augmentation research, may not capture the full trajectory of skill drift: some theoretical mechanisms — particularly procedural memory atrophy — may produce their most consequential effects over multi-year timeframes. Future research employing 12-to-24-month longitudinal designs is strongly warranted. The sample, while spanning three knowledge work contexts, is limited to relatively junior workers (< 5

years mean experience); the skill drift implications of AI copilot use for mid-career and senior experts — whose expertise is both more deeply consolidated and potentially more resilient to atrophy — require dedicated investigation.

The study also does not examine skill drift recovery: whether and how quickly independent capabilities can be rehabilitated following skill drift through deliberate practice reintroduction is a theoretically important and practically urgent question that future experimental research should address. Finally, the adaptive AI condition (C4), while demonstrating superior outcomes, was implemented through a purpose-built research intervention; whether commercial AI copilot systems can be designed to achieve equivalent adaptiveness at scale is an engineering and product design challenge that requires substantial future development and evaluation.

VI. CONCLUSION

The question at the heart of this study — do AI copilots enhance expertise or create cognitive dependency? — resists a simple answer, and that complexity is itself the study's most important finding. AI copilots are not inherently upskilling or dependency-creating; they are designed interventions whose long-term human capability consequences depend critically on the intensity, adaptiveness, and governance context of their deployment. Static high-intensity AI assistance generates the productivity improvements that dominate current AI research while simultaneously producing skill drift that those productivity metrics are structurally incapable of detecting. Adaptive AI assistance, by contrast, largely dissolves this trade-off — delivering comparable productivity benefits while preserving the independent skill profiles that constitute long-term human capital.

The practical stakes of these findings are considerable. Organizations that deploy AI copilots at maximum intensity to maximize short-term productivity metrics are, this study suggests, making a long-term human capital investment decision of which they are largely unaware. The workforce that produces higher-quality, faster outputs today — because AI is generating most of the cognitive content — may be a less capable workforce in two years, one that is faster and more productive when AI-assisted but more fragile, less accurate, and less reliable when AI is unavailable, degraded, or incorrect. In a world of increasingly capable but imperfect AI, the ability to evaluate, verify, and override AI outputs is a critical organizational capability. Building that capability — rather than inadvertently eroding it — should be a central design criterion for responsible AI copilot deployment.

The path forward is not to restrict AI copilot use but to design it with long-term human capability as an explicit optimization target alongside productivity and quality. The Responsible Augmentation Framework proposed in this study provides a starting conceptual architecture for that design challenge. The future of human-AI collaboration is not one in which humans defer entirely to AI or maintain AI-free independence; it is one in which AI systems are designed to make humans genuinely more capable — not merely more productive — over time. That future requires both theoretical clarity about the mechanisms of skill drift and practical commitment to the governance and design choices through which responsible augmentation can be achieved.

REFERENCES

1. Anderson, J. R. (1983). The architecture of cognition. Harvard University Press.
2. Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3–30.
3. Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8
4. Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
5. Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. National Academy Press.
6. Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER Working Paper No.

31161. National Bureau of Economic Research. https://doi.org/10.3386/w31161
7. Carr, N. (2014). The glass cage: Automation and us. W. W. Norton & Company.
8. Charness, N., & Tuffiash, M. (2008). The role of expertise research and human factors in capturing, explaining, and producing superior performance. Human Factors, 50(3), 427–432.
9. Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). The nature of expertise. Lawrence Erlbaum Associates.
10. Clark, A. (2003). Natural-born cyborgs: Minds, technologies, and the future of human intelligence. Oxford University Press.
11. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems. Proceedings of the 52nd Hawaii International Conference on System Sciences.
12. Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100(3), 363–406.
13. Ericsson, K. A., & Pool, R. (2016). Peak: Secrets from the new science of expertise. Eamon Dolan/Houghton Mifflin Harcourt.
14. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906–911.
15. Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280.
16. Gentner, D., & Stevens, A. L. (Eds.). (1983). Mental models. Lawrence Erlbaum Associates.
17. Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.
18. Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. Human Factors, 53(5), 517–527.
19. Jagatha, A., Sammangi, H., & Maddireddy, H. G. (2025). Decentralized multi-hop federated reinforcement learning for energy-efficient and secure routing in LoRaWAN-based smart city infrastructure. Engineering: Open Access, 3(6), 1–8.
20. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
21. Kane, G. C., Phillips, A. N., Copulsky, J. R., & Andrus, G. R. (2019). The technology fallacy: How people are the real key to digital transformation. MIT Press.
22. Kohnert, A., & Haenel, T. (2024). Cognitive dependency in AI-assisted work: A longitudinal examination. Computers in Human Behavior, 158, 108301.
23. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology, 77(6), 1121–1134.
24. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.
25. Licklider, J. C. R. (1960). Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, HFE-1(1), 4–11.
26. Luria, A. R. (1976). Cognitive development: Its cultural and social foundations. Harvard University Press.
27. McKinsey Global Institute. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.
28. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586
29. Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. Organization Science, 3(3), 398–427.
30. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics Part A, 30(3), 286–297.
31. Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. Human Factors, 50(3), 511–520.
32. Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact

assessments: A practical framework for public agency accountability. AI Now Institute.
33. Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100(3), 349–363.
34. Sammangi, H., Ambati, L. S., Liu, J., & Jagatha, A. (2025). AI-driven decentralized IoT for secure and scalable healthcare. AMCIS 2025 Proceedings. https://aisel.aisnet.org/amcis2025/health_it/sig_health/3
35. Sammangi, H., Jagatha, A., & Liu, J. (2025b). Harnessing generative AI and large language models for revolutionizing cybersecurity in the Internet of Things: Ethical and privacy implications. Engineering: Open Access, 3(6), 1–12.
36. Sammangi, H., Jagatha, A., & Liu, J. (2025c). Integrating blockchain technology into telemedicine: A framework for enhancing data privacy and security. Engineering: Open Access, 3(6), 1–7.
37. Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333(6043), 776–778.
38. Sharma, G., Singh, J., Sammangi, H., Sharma, M., Pandey, R., Srivastava, S., Agarwal, G., & Singh, I. (2025a). A comprehensive assessment of developing a forecasting model for kidney stone formation using deep learning approaches. In H. Sharma, A. Chakravorty, S. Hussain, & R. Kumari (Eds.), Artificial Intelligence: Theory and Applications (Vol. 5588, pp. 121–132). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1918-4_9
39. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.
40. Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Knopf.
41. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.
42. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
43. World Economic Forum. (2025). The future of jobs report 2025: AI, automation and workforce transformation. WEF.
44. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.