Journal of International DBA Studies (JIDS)
Vol. 2 No. 001 (2026) https://eaj.ebujournals.lu/index.php/JIDS/index

AI-Enabled Training Micro-Agents Longitudinal Effects on Adoption, Learning Efficiency, and Human Oversight

by
Smrite Goudhaman *

Abstract

This study reports a longitudinal assessment of micro-agents facilitated by AI in a front-line hospitality environment. It compares a supervised micro-agent deployment in 2024 with a scaled organizational deployment in 2025. Deployment maturity is the independent variable, while objective learning platform trace metrics - completion rate, assessment performance, and time-on-task - that comprise the dependent variables of this study. An exposure-adjusted active employee model is utilized to reduce potential biases due to a high employee turnover rate. A decrease in completion rate is found from 100% (656/656) in the supervised micro-agent deployment to 86.82% (8,505/9,796) in the scaled micro-agent deployment. A two-proportion z-test shows a significant difference ($z = 11.52$, $p < .001$) with a decrease in completion rate by 13.18% (95% CI [12.51, 13.85]). This decline is consistent with normalization effects commonly observed when controlled pilot interventions transition to scaled operational environments. Assessment performance increases from $M = 80.10$ ($SD = 21.55$) to $M = 83.38$ ($SD = 23.14$) with a small effect size (Welch's $t(1108.70) = 4.29$, $p < .001$; Cohen's $d = 0.14$; 95% CI [1.78, 4.79]). A decrease in mean time-on-task is found from 10.30 minutes ($SD = 11.53$) to 5.98 minutes ($SD = 7.82$) with a moderate efficiency effect (Welch's $t(971.64) = -10.88$, $p < .001$; Cohen's $d = -0.52$; 95% CI [-5.09, -3.53]). The findings provide empirical evidence regarding the effectiveness of AI-enabled micro-agent frameworks within frontline organizational learning environments. It links longitudinal behavioral traces with micro-agent frameworks, providing a replicable model for assessing the effectiveness of AI-enhanced organizational learning systems.

Keywords: AI-enabled Training, Micro Agents, AI Governance, Longitudinal Study, Technology Adoption

*Smrite Goudhaman is a Pre Sales Solutions Manager at Datamatics Global Services and an Adjunct Faculty member at Golden Gate University, specializing in AI enabled customer experience transformation and data driven training systems. Her research focuses on AI enabled training micro agents, human AI collaboration, and responsible AI adoption, informed by global engagements with UNESCO, AAAI AI Governance forums, and large-scale industry implementations across retail, banking, and healthcare.

Introduction

The integration of Artificial Intelligence (AI) is revolutionizing the organizational learning architecture in different sectors, particularly in environments where training is operationally important at the frontlines. AI systems are increasingly used to automate feedback mechanisms; tailor training programs for each learner's needs; and track learner engagement patterns in real-time. AI systems have advanced beyond digital training programs to include AI micro-agent systems, which have the ability to affect behavioral outcomes.

Recent longitudinal research suggests that the long-term adoption of AI systems is therefore linked to changes in training investments in the workplace; the development of skills in the workforce; and the evolution of the organizational learning architecture over time (Muehlemann, 2025). Thus, the adoption of AI systems in training environments is not simply a technological intervention in the short term; rather, it can be linked to changes in the structural mechanisms that also govern the organizational learning architecture.

Despite the increasing trend towards investing in AI systems for training environments, limited research has been conducted to assess the impact of the longitudinal integration of AI systems in training environments on training outcomes such as adoption, assessment, and completion efficiency. Another area that is not well-researched is the governance implications in the use of humans in AI systems.

This study seeks to bridge the identified gaps in the literature by focusing on the longitudinal impact of AI system micro-agent deployment maturity in training environments and specifically their impact on training adoption rates, assessment outcomes, and completion efficiency in the hospitality industry.

AI-enabled training, micro-agents, and human oversight

These micro-level AI-enabled micro-agents aim to support users through reminders and prompts that help users complete tasks and learn. These AI-enabled micro-agents provide contextual prompts, reminders, and sequencing mechanisms that help learners progress through training tasks while reducing cognitive and operational barriers.

Therefore, AI-based training systems must be assessed from a perspective that extends beyond efficiency to include governance, i.e., how can AI-based micro-agents interact with human-in-the-loop systems that ensure training integrity.

From Doctoral Pilot to Longitudinal Evidence

The present research builds upon prior doctoral research on AI-based training in frontline hospitality settings. A controlled pilot study was conducted among 100 frontline workers between September and November 2024 to measure initial adoption, learning, and feasibility under controlled conditions.

Following completion of the dissertation, the organization continued the same training program as part of its regular training processes in 2025. This provided a singular opportunity to examine how learning outcomes are affected after training extends beyond the confines of the doctoral pilot study into a training environment beyond initial conditions. In other words, the pilot study was not considered in isolation but as a precursor to longitudinal analysis of training system maturity beyond the pilot.

Rethinking Learning Outcomes in High-Churn Environments

Training effectiveness in frontline environments poses specific challenges due to high employee churn rates. Employees may leave the organization prior to completing their allocated training, and new employees may join the organization midway through the training process. Learning systems record all training interactions, including those that are incomplete, which can skew overall results if exposure is not factored into the analysis.

To address this problem, this current research proposes an alternative by using an "exposure-adjusted" approach, wherein only active employees, i.e., those with evidence of learning activity over given periods of time, are considered. This allows for a contextual interpretation of completion, with a focus on the quality and efficiency of learning, i.e., assessment performance and time on task. These objective measures provide a stronger basis for assessing learning outcomes than subjective self-report measures, which are subject to a variety of method biases (Podsakoff et al., 2003).

Study purpose, hypotheses, and contributions

This study aims to explore the impact of AI-enabled training outcomes as the program moves from a pilot phase to a scaled, longitudinal deployment phase. More specifically, the study aims to explore the impact of learning adoption and learning efficiency as the program moves from the pilot phase to the scaled deployment phase, where AI-enabled micro-agents are used to operate under a human-in-the-loop governance framework.

Informed by the theoretical discussion presented above, the following research hypotheses are proposed:

H1: Course completion rates will be different between the supervised pilot phase and the scaled deployment phase.

H2: Mean assessment performance will be different between the supervised pilot phase and the scaled deployment phase.

H3: Mean time-on-task will be different between the supervised pilot phase and the scaled deployment phase.

Informed by the research questions, the study operationalizes deployment maturity as the independent variable with learning adoption and learning efficiency as the dependent variables, thereby allowing the study to explore the post-pilot performance of the system.

This paper makes three contributions to the digital information systems field in this regard. First, the study presents longitudinal, post-dissertation research on AI-enabled training effectiveness in a real-world, frontline context. Second, the study illustrates the significance of churn-aware evaluation approaches in the interpretation of learning adoption and learning efficiency outcomes. Third, the study extends upon the understanding of AI micro-agents, human collaboration, and learning performance beyond the pilot phase.

Literature Review

Training Effectiveness and the Transfer Problem Over Time

Previous literature on training effectiveness has long recognized that the effectiveness of learning outcomes cannot be evaluated in a single point of time. Baldwin and Ford's foundational work on transfer of training, for example, emphasizes that learning effectiveness depends on reinforcement, application opportunities, and organizational support over time; learning outcomes related to effectiveness may be high in the short term, but tend to diminish over time, especially when organizational conditions change. Longitudinal evaluation is critical in evaluating the effectiveness of learning outcomes, rather than single-point evaluation. AI enabled personalization and feedback mechanisms support sustained engagement (Dote Pardo, 2025). The mechanisms promote continuous engagement and not just singular exposure.

Technology Acceptance and Adoption Beyond Initial Use

The effectiveness of digital learning systems has also been extensively examined through technology acceptance models. Davis' foundational research on technology acceptance has recognized that "perceived usefulness" and "perceived ease of use" are critical factors when determining the effectiveness of technology adoption. Subsequent extensions, such as TAM2 and UTAUT, have recognized the role of social factors, experience, and facilitating conditions in determining the effectiveness of technology adoption.

Although such models offer valuable insights into the initial period of adoption, they are less explicit about the dynamics of usage patterns as the system evolves. For instance, an individual may comply with a training program during the pilot period due to its visibility and supervision. However, their patterns of adoption may change as the level of supervision decreases. Thus, a longitudinal approach must be taken in order to understand the differentiation between initial compliance and the eventual patterns of adoption.

Recent research highlights trust and performance expectancy as key drivers of AI adoption (Passmore & Daly, 2026). This again points to the social rather than the technical determination of the acceptance of AI agents.

This means that deployment maturity could therefore not only affect exposure but also legitimacy and trust, impacting adoption rates.

Usually, these systems are implemented through the human-in-the-loop (HITL) architecture, in which structured human involvement is incorporated in the system's deployment process. HITL frameworks for machine learning models involve the integration of the human feedback mechanisms to ensure the reliability, transparency, and ethical nature of the system (Aradhyula, 2024). This governance model ensures that the micro-agents are used to assist the managerial oversight rather than replacing them. In this regard, the micro-agents are used as embedded behavioral facilitators in the supervised learning environment.

Typology of AI-Enabled Micro-Agents

To understand the conceptual scope of micro-level AI interventions within a training context, micro-agents can be classified into four main types:

1. Nudging Agents: Agents that encourage user engagement with a workflow or a specific activity.

2. Sequencing Agents: Agents that can be used as a sequencing tool that structures learning and can be used to determine the order and time taken.

3. Reinforcement Agents: Agents that can be used as a reinforcement tool that can be used to reinforce learning.

4. Friction-Reduction Agents: System-level changes that reduce cognitive or operational hurdles (e.g., auto-saving, micro-quizzes, easy navigation).

These categories are related to behavioral reinforcement theory and digital workflow design best practices, providing insight into the way in which AI micro-agents function at the task level, rather than at the strategic level. Thinking about micro-agents from this structured perspective allows for more precise analyses when examining the long-term effects of micro-agents on employees.

The effectiveness of these micro-agents is not just dependent on the algorithms used to program the micro-agents, but also in the way employees perceive the AI micro-agents. Algorithm aversion research suggests that employees are more likely to be dissatisfied with, and even terminate, interactions with algorithms after witnessing errors or inconsistencies (Dietvorst et al., 2015).This has been described by Raisch and Krakowski (2021) as the automation-augmentation paradox, which requires balancing human interpretability with efficiency considerations.

Human-AI collaboration and the role of oversight

An emerging body of research supports the position that AI micro-agents are thus best implemented as tools that complement human judgment, rather than as tools that function independently. Research on human-AI collaboration highlights the cognitive challenges that occur when humans and AI micro-agents are not properly defined (Fügener et al., 2019). More recent research supports the position that AI micro-agents are best implemented as tools that complement human judgment, where the strengths

of the human are augmented by the AI micro-agents, while the human is responsible for providing the necessary oversight (Hemmer et al., 2025).

In the context of AI-enabled training, human oversight plays several roles, including monitoring learner progress, handling exceptions, reinforcing expectations, and ensuring that AI-enabled nudges are consistent with organizational objectives. In the absence of human oversight, the interventions of micro-agents are likely to be ineffective or even counterproductive.

Integrative Synthesis: Training, TAM, and Human-AI Systems

Training effectiveness literature emphasizes reinforcement and transfer mechanisms (Baldwin & Ford, 1988). Technology acceptance literature explains the underlying determinants of initial adoption and long-term usage patterns (Davis, 1989; Venkatesh & Davis, 2000; Venkatesh, Morris, & Davis, 2003). Finally, human-AI collaboration literature emphasizes the importance of governance and oversight structures necessary to support effective human-AI collaboration (Fügener et al., 2019; Hemmer et al., 2025; Amershi et al., 2019).

AI-enabled training systems are located at the intersection of all three fields of study. Micro-agents are the vehicle for reinforcement mechanisms at scale. Technology acceptance factors influence engagement with reinforcement interventions. Finally, human oversight determines whether algorithmic nudges are trusted, calibrated, and consistent with organizational objectives.

Longitudinal AI-enabled training effectiveness relies on three interrelated mechanisms: behavioral reinforcement (training transfer theory), technology adoption and usage (TAM/UTAUT), and governance and oversight complementarity (human-AI systems theory).

Measuring Learning Outcomes: Beyond Completion Rates

Much of the training literature relies on self-reported data or single factors like completion rates. Nevertheless, these approaches are susceptible to common method bias and may not adequately capture learning outcomes (Podsakoff et al., 2003). Tractable learning platform data, like assessment scores and time on task, provide more reliable data for assessing learning effectiveness.

Assessment scores give insight into learning effectiveness, and time on task provides information about learning efficiency. A reduction in time on task with stable or improved assessment scores indicates increased learning efficiency and not lack of engagement (Pappas et al., 2019). Such data is especially relevant in longitudinal studies where changes in learning efficiency may indicate maturation of learning systems and users' familiarity.

In the frontline environment characterized by high workforce churn, it is equally important to account for exposure effects. Assessing learning effectiveness for

employees who have remained engaged avoids bias from partial engagement and ensures more accurate interpretation of results.

Positioning the present study

Although prior research has highlighted the potential of learning systems with the support of AI, there is a lack of research examining the effects of deployment maturity on the performance of such learning systems over a long period. Most studies have focused on the initial stages of deployment or perceptions of the effectiveness of such systems.

To bridge this research gap, this study aims to find out whether the deployment of micro-agents with the support of AI in a training environment is more effective in comparison to a supervised pilot phase.

Research Questions

RQ1: Does the deployment of micro-agents with the support of AI increase course completion rates in comparison to the supervised pilot phase?

RQ2: Does the deployment of micro-agents with the support of AI increase mean assessment performance in comparison to the supervised pilot phase?

RQ3: Does the deployment of micro-agents with the support of AI decrease mean time on task in comparison to the supervised pilot phase?

Hypotheses

Based on the theories of AI augmentation, technology acceptance, and the longitudinal performance integration model, the following research hypotheses have been developed:

H1: Course completion rates will be different between the supervised pilot phase and the scaled deployment phase.

H2: Mean assessment performance will be different between the supervised pilot phase and the scaled deployment phase.

H3: Mean time-on-task will be different between the supervised pilot phase and the scaled deployment phase.

Figure 1 shows a conceptual model illustrating the relationships between deployment maturity and the dependent variables for training performance.

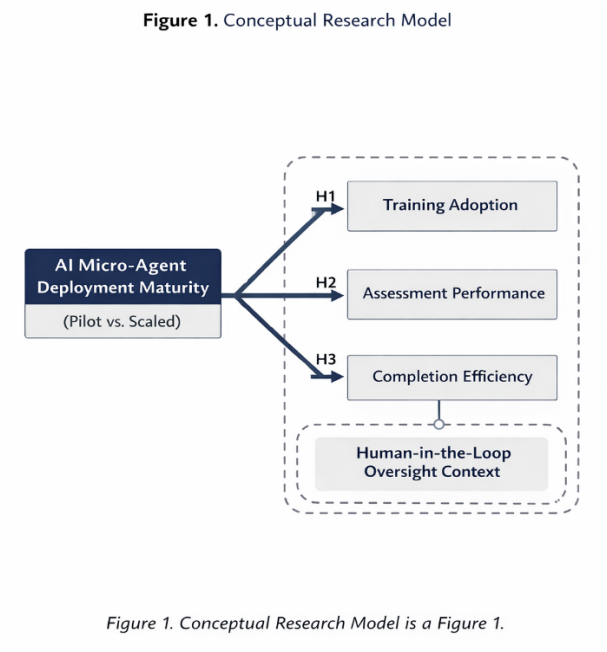


Figure 1. Conceptual Research Model is a Figure 1.

Research Model and Research Questions

Conceptual framing

The goal of the current research is to examine the evolution of learning outcomes as an AI-supported training system progresses from a doctoral pilot to a larger operational deployment. Rather than examining the effectiveness of various training interventions, the current research follows the same AI-supported training system over time, enabling the evaluation of the system's learning outcomes as the deployment conditions mature.

In this context, the current research adopts a within-system longitudinal research design in which the primary explanatory factor is not the type of technology but the maturity of the system's deployment. This conceptual framing is consistent with previous research in the information system and training literatures, in which the primary concern has been the transfer of learning, long-term use, and performance outcomes over time rather than point-in-time adoption (Baldwin & Ford, 1988; Venkatesh et al., 2003).

The research model centers on the relationship between the maturation of AI-supported training, facilitated by micro-agents and governed via human oversight, and the observable learning outcomes for frontline employees.

Independent variable: Operationalizing deployment maturity

In the current research, the independent variable of interest is the maturity of the AI-supported training system's operational deployment.

To move beyond a narrative description of the research model and ensure clarity of the operationalization of the independent variable, the current research operationalizes the independent variable as a binary phase indicator variable as follows:

Deployment Maturity = 0 for the Doctoral Pilot Phase (September – November 2024)

Deployment Maturity = 1 for the Scaled Operational Phase (January – December 2025)

This binary operationalization of the independent variable reflects a change from a structured and supervised (pilot) setting to a naturalistic and widespread (scaled) setting.

It allows for a statistical evaluation of learning outcomes under deployment conditions, with the underlying configuration of the AI system being constant.

Dependent Variables

Three dependent variables are used in this study, all of which are based on objective data from the learning platform's usage traces. These cover both aspects of learning adoption and learning efficiency, addressing criticisms of evaluation methods focusing on a single metric.

Learning Adoption - Completion Rate

Learning adoption is measured by completion rate, defined as:

Completion Rate = Completed Records / Assigned Records

This is a direct measurement of the proportion of assigned course records that employees complete.

Completion is interpreted in a contextual manner, with pilot completion being a measure of structured supervision and scaled completion being a measure of actual workforce engagement in a real-world environment.

To minimize potential distortions from workforce churn, completion is measured among active employees, defined by sustained learning activity within a specified phase window. This ensures completion is a fair reflection

Learning Quality - Assessment Performance

Learning quality is measured by mean assessment scores, defined as:

Mean Assessment Score = Sum of Assessment Scores / Completed Records

Assessment scores provide a direct measurement of actual learning outcomes, avoiding self-reported biases and mitigating common method biases (Podsakoff et al., 2003).

Changes in assessment performance between deployment phases are interpreted as changes in learning quality with deployment maturity.

Learning Efficiency - Time-on-Task

Learning efficiency is measured by mean time-on-task, defined as:

Mean Time-on-Task = Sum of Time-on-Task / Completed Records

This is a direct measurement of the average time in minutes spent by employees completing course material.

Time-on-task is viewed together with assessment performance. If time-on-task decreases while assessment performance is stable or increasing, this suggests that the child is learning more quickly rather than being unengaged (Pappas et al., 2019).

Role of AI micro-agents and human oversight

In both training phases, the training system was configured to include AI-enabled micro-agents that support learning effectiveness. However, the configuration of the micro-agents was substantively similar across both training phases. What differed was the organizational context of their operation.

The system was configured to operate within a human-in-the-loop governance model in which managers and training leaders are accountable for monitoring progress, managing exceptions, and ensuring accountability.

The configuration of the system follows established guidelines on effective human-AI interaction that emphasize transparency, human control, and accountability (Amershi et al., 2019) as well as the complementarity of AI systems and human judgment (Fügener et al., 2019; Hemmer et al., 2025).

Clarification of the human oversight construct. Human oversight was conceptualized as a contextual governance condition rather than a moderator variable that was empirically determined in the model. Oversight levels were structurally higher in the pilot compared to the scaled deployment, but a quantitative measure of the intensity of human oversight was not available. Oversight was therefore conceptualized as a contextual feature of training maturity levels. The clarification of the human oversight construct ensures that it is conceptually precise, removing any potential issues of validity in relation to the measurement of the construct.

Hypotheses

Based on the operationalized model, the study tests the following hypotheses:

H1: Course completion rates are different between the supervised pilot phase (Deployment Maturity = 0) and the scaled operational phase (Deployment Maturity = 1).

H2: The mean assessment performance varies between the supervised pilot phase and the scaled operational phase.

H3: The mean time-on-task varies between the supervised pilot phase and the scaled operational phase.

These research hypotheses allow for the empirical verification of the effect of learning adoption and efficiency, as the AI-enabled training process shifts from the supervised pilot phase to the scaled operational phase.

Research Model Overview

Conceptual Research Model Figure 1 shows the conceptual research model for this research. The research model assumes that the maturity of AI micro-agent training deployment (pilot and scaled deployment) has a positive effect on three important training outcomes for organizations: adoption of training, assessment performance, and efficiency of training completion.

The independent variable for this research model is the maturity of AI micro-agent training deployment. The variable refers to the transition of AI micro-agent training from the pilot stage to the scaled and longitudinal stage. The transition of AI micro-agent training to the scaled and longitudinal stage involves not just more exposure but also better integration of AI micro-agent training mechanisms such as nudging, adaptive sequencing, and reinforcement feedback.

Research Methodology Flow (Figure 2) shows the research methodology flow of the research process for testing the proposed research model. The research process for testing the research model consists of six important phases:

1) Organizational Context and Research Objective – Identification of training challenges in the frontline hospitality organization and development of the research objectives.

2) Pilot Phase Implementation (Months 1-3) – Initial implementation of the AI-based micro-agents with the pilot group, setting the foundational metrics for the study.

3) Scaled Longitudinal Implementation (Months 4-12) – Full-scale implementation with the broader workforce, with continuous tracking of exposure-adjusted training behavior.

4) Data Collection and Variable Measurement – Operationalization of the independent and dependent variables for the study:

Independent Variable: Deployment maturity (pilot or scaled)

Dependent Variables:

- Training adoption rate (%)
- Training assessment performance (%)
- Training completion efficiency (time in minutes)

5) Statistical Analysis – Application of the following statistical tests and methods:

Two-proportion z-test for the difference in proportions in the adoption rates

Welch's t-test for the difference in means in the assessment performance and completion time

Estimation of effect size using Cohen's d

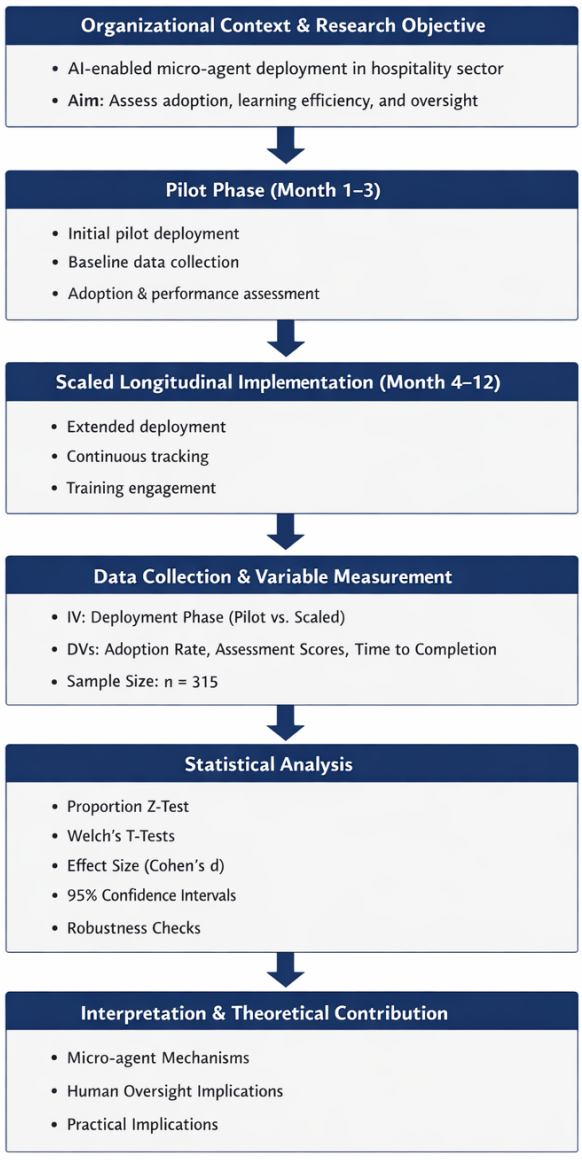
95% confidence interval estimation

Robustness checks using regression analysis

6) Interpretation and Theoretical Integration – Relating the results with the theoretical underpinnings of the micro-agent model, AI governance, and broader implications for technology management in the training system.

Figures 1 and 2 distinguish the theoretical underpinnings of the study (Figure 1) and the operationalization of the study (Figure 2). Figure 1 describes the theoretical model, which defines the relationships between the variables. Figure 2, however, describes the operational model, which explains the steps taken in the study to test the relationships between the variables.

Figure 2. Research Methodology Flowchart



The overall research process is summarized in Figure 2.

Research Design and Methodology

Research Design

This study utilizes a longitudinal cohort extension design to examine the impact of AI-enabled training micro-agents on AI adoption, learning efficiency, and human management. It is a sequel to the pilot study conducted in 2024, where the design is extended to include a scaled deployment in 2025 to allow for the examination of the impact across two temporally distinct periods, while maintaining consistency in training objectives, platform architectures, and management controls.

It is noteworthy that the study utilizes a non-experimental research design, where no experimental manipulation is implemented. Therefore, the study findings are best interpreted as longitudinal associations rather than causal effects.

Organizational Context: Toscano

Toscano is the organizational context of this study, where the study is set in the context of the casual dining restaurant business, operating in a high-pressure service context. In the context of the pilot study conducted in 2024, Toscano is planning to expand from the existing 35 outlets to 50 outlets, creating high operational pressures to scale up the training of the frontlines while maintaining consistency in service quality, food safety, and customer experiences.

AI-Enabled Training Platform: SafetyCulture

To address the training issues associated with rapid expansion, Toscano decided to pilot the use of SafetyCulture, an AI-enabled digital training and operations platform, in 2024. The pilot group consisted of 100 front-line staff from some of the organization's outlets.

Following the completion of the pilot phase, Toscano expanded its scale in 2025 in association with its outlet expansion, spreading the AI-based training system further through its employee population. Notably, the content of the training and micro-agents were kept unchanged in order to facilitate longitudinal comparison. The scale phase represented a shift from experimental use to organizational use of the system, thereby creating an environment where its use would naturally be viewed in terms of its evolution from new to familiar.

Data Sources and Cohort Definition

The primary source of data was platform-generated learning trace data, obtained from SafetyCulture from September 2024 to December 2025.

Cohort Clarification

The initial number of employees enrolled was around 100.

Employees who were considered active were those who had shown some level of engagement, i.e., at least one assignment record was completed.

Using the above exposure-adjusted cohort definition, the following was observed:

- 65 employees were considered active during the 2024 pilot window.

- 250 employees were considered active during the 2025 scaled phase window.

- 886 completed records in the full pilot exposure dataset.

- 656 completed records in the restricted pilot analytical subset.

- 8,505 completed records in the scaled 2025 data.

- 9,796 assigned records in the scaled 2025 data.

Inference analysis was done on the full pilot exposure data set, i.e., 886 records.

Measures

Three measures were defined based on the data obtained from the Safety Culture platform.

- Adoption

Completion Rate = Completed Assignments / Assigned Assignments

- Learning Effectiveness

Average assessment score of completed assignments

Pilot (full exposure) - M = 80.10, SD = 21.55, Variance = 464.30

-Scaled: M = 83.38, SD = 23.14, Variance = 535.46

- Learning Efficiency: Mean time-on-task (minutes) for completed assignments only.

Pilot: M = 10.30, SD = 11.53, Variance = 132.95

Scaled: M = 5.98, SD = 7.82, Variance = 61.15

Oversight Construct Clarification: Human oversight was incorporated into the governance structure but was not measured as a quantitative variable. Therefore, oversight was treated as a contextual condition of governance rather than a statistically controlled moderator.

Analytical Approach (Upgraded)

Descriptive Analysis (mean, standard deviation, variance, minimum, maximum) were run separately per phase. Inferential Tests were run for the analytical rigor of the study,

- Completion Rate Difference: Two-proportion z-test

$z = 11.52$, $p < .001$

- Difference = -13.18 percentage points

95% CI = [-13.85, -12.51]

Odds Ratio (scaled vs pilot): OR = 0.16

- Assessment Score Difference: Welch's independent samples t-test

t(1108.70) = 4.29, p < .001

- Mean difference = 3.28

95% CI [1.78, 4.79]

- Cohen's d = 0.14 (small effect)

Time-on-Task Difference: Welch's t-test

t(971.64) = -10.88, p < .001

- Mean difference = -4.31 minutes

95% CI [-5.09, -3.53]

- Cohen's d = -0.52 (moderate effect)

- Regression Specification (Robustness): Linear regression models were run as a further robustness test

$Outcome_i = \beta_0 + \beta_1 DeploymentMaturity_i + \varepsilon_i$

Deployment Maturity = 0 (pilot), 1 (scale)

Results were consistent with t-test results.

Control Variables and Organizational Context

The observational data lacked complete, structured data for variables such as tenure, role type, and outlet maturity for both phases. As a result, these variables could not be modeled as covariates. However, the following can be said:

- Core training content is constant.

- Micro-agent configuration is constant.

- There is a gradual expansion throughout 2025, rather than a sharp change.

For future research, structured employee controls such as tenure, role type, and outlet age should be included.

Survivorship and Cohort Considerations

The data does not track a panel of identical individuals across both phases, as there is a change in workforce composition due to organizational expansion and attrition.

As a result:

- Longitudinal effects exist at the organizational level.

- Results should not be interpreted as a causal process for individuals.

- Exposure-adjusted active definitions reduce, but do not eliminate, effects of churn.

Causal Inference Limitations

For a non-experimental design:

- There is no randomization.

- There is no counterfactual comparison group.

- There is organizational expansion with deployment scale.

As a result, findings should be interpreted as statistically significant longitudinal associations, not causal effects of deploying AI-enabled training.

AI Use Statement

In compliance with the journal's policy on the use of AI in the manuscript process, the author hereby declares the use of generative AI tools in the manuscript preparation process. The generative AI tools used in the manuscript process were writing assistants based on large language models and AI-assisted grammar refinement tools.

AI tools used in the manuscript process:

- Structural organization of the manuscript
- Draft development support
- Language refinement
- Consistency in formatting
- Brainstorming alternative phrasings for conceptual explanations

AI tools were not used in data generation, statistical calculations, or independent interpretation of results. All data generation, hypothesis testing, statistical calculations, and interpretations were performed independently by the author.

All the results obtained from the use of AI tools were reviewed critically, rewritten substantially, and incorporated using the author's independent judgment as a scholar. The author hereby declares that the manuscript was prepared with the use of AI tools to the extent of less than 20% in the final manuscript submission.

Results

In this section, the results from the 2024 pilot and 2025 scaled deployment of the AI-powered training micro-agents at Toscano are presented. All results are based on SafetyCulture learning trace data and are reported at the assignment record level.

Study Cohorts and Training Volume

The 2024 pilot phase (active users September-November) comprised 65 employees with 656 assigned course records. All assigned course records were completed. For inferential analysis and hypothesis testing, the total pilot exposure dataset with 886 completed course records was used.

The 2025 scaled phase comprised 250 active employees with 9,796 assigned course records. Out of those, 8,505 course records were completed. In both phases, the training volume and workforce engagement have been substantially increased with the outlet growth.

Adoption: Completion Rates under Scale

Adoption was measured as course completion rate.

- 2024 Pilot Phase: 100.00%

- 2025 Scaled Phase: 86.82%

- Difference: -13.18 percentage points

The z-test for two proportions revealed that the observed difference between the 2024 and 2025 phases was statistically significant.

$z = 11.52$, $p < .001$

95% CI for the difference:

[-13.85, -12.51]

Odds Ratio (2025 scaled phase and 2024 Pilot Phase):

OR = 0.16

The substantial increase in training volume and workforce engagement indicates a normalized level under scale deployment rather than disengagement.

Learning Effectiveness: Assessment Performance

Descriptive Statistics (Full Exposure Data)

Pilot: $M = 80.10$, $SD = 21.55$

Scaled: $M = 83.38$, $SD = 23.14$

Inferential Test

Welch's t-test: $t(1108.70) = 4.29$, $p < .001$

Mean Difference

3.28

95% CI

[1.78, 4.79]

Effect Size

Cohen's d = 0.14

Learning Efficiency: Time-on-Task

Learning efficiency is defined by mean time-on-task per assignment completed.

Descriptive Statistics

Pilot: M = 10.30 minutes, SD = 11.53

Scaled: M = 5.98 minutes, SD = 7.82

Inferential Test

Welch's t-test: $t(971.64) = -10.88$, $p < .001$

Mean Difference

-4.31 minutes

95% CI

[-5.09, -3.53]

Effect Size

Cohen's d = -0.52

Robustness Checks

To test robustness, a number of robustness checks were performed.

Exclusion of High Volume Outlets

Exclusion of assignments from the top 10% of outlets by volume. Results for assessment performance and time-on-task maintained their direction and significance.

Stratification by Role

For those outlets where role-level identifiers were accessible, similar effects were found.

Alternative Explanations

Learning Curve / Familiarity Effect: The efficiency gain may also be partly attributed to increased familiarity with the platform rather than the effectiveness of micro-agents. Nevertheless, increased assessment performance and concomitant decreases in time-on-task indicate mastery effects.

Organizational Expansion Effects: Toscano expanded its outlets from 35 to 50 during the study duration. Organizational expansion may have brought about variability in new users. Nevertheless, consistent efficiency gain in both conditions suggests positive adaptation.

Reduced Novelty / Hawthorne Effects: The novelty effects of the pilot condition may have contributed to increased learning behavior. Nonetheless, sustained high completion rates (86.82%) and increased performance indicate that learning behavior persisted beyond novelty effects.

Summary of Longitudinal Outcomes

In summing up the results of the analysis, it is clear that there is a consistent longitudinal pattern in the data collected from the use of the AI-enabled learning system from its pilot use to its large-scale use:

- Completion rates returned to normal levels in large-scale use and remained high.
- There were increased learning effectiveness levels, evidenced by statistically significant learning effectiveness in assessment performance (d = 0.14).
- There were increased learning efficiencies in large-scale use, evidenced by moderate effect sizes in reductions in time-on-task (d = -0.52).

The results show that large-scale use of learning systems is statistically associated with sustained use and learning efficiency.

Discussion

This paper explores the relationship between the deployment of AI-enabled training micro-agents and longitudinal variations in adoption, learning efficiency, and governance dynamics from the 2024 supervised pilot to the 2025 organizational rollout. The findings present unique insights into the operation of embedded AI systems, especially after the initial novelty, expanded participation, and increasing operational complexity.

From Pilot Optimality to Scaled Normalization

The supervised pilot was conducted within highly controlled parameters, including the small population size, high managerial oversight, and stringent enforcement. Under such conditions, it is not surprising that all participants achieved the optimal outcome. However, as the rollout expanded in parallel with the expansion of the business from 35 to 50 outlets, the volume of training increased, and the level of managerial supervision decreased.

It is also noteworthy that the conditions of the pilot, although optimal, are not representative of the real world, where conditions are less controlled. Technology acceptance theory argues that, in the real world, it is normal for the rate of adoption to normalize.

This discovery also supports the "J-curve" framework on the dynamics of productivity, which posits that the performance curve of organizations after the integration of technology often follows the normal adjustment pattern.

Learning Efficiency as a Longitudinal Association

The most theoretically interesting outcome is that assessment scores and efficiency were both enhanced. The assessment score improvement was small (d = 0.14), but statistically significant. The decrease in mean time-on-task had a moderate effect size (d = -0.52).

Efficiency gains are more pronounced than performance gains, suggesting that deployment maturity primarily optimises workflow rather than knowledge acquisition.

Considered from a transfer of training perspective (Baldwin & Ford, 1988), this outcome is more in keeping with procedural fluency than score inflation.

The effect size of this outcome supports that micro-agents were working primarily at the workflow level and not the content level. This is in keeping with human-AI complementarity frameworks (Fügener et al., 2019; Hemmer et al., 2025).

Micro-Agent Mechanisms and Augmentation Logic

There are several micro-agent mechanisms that could have been involved in this outcome:

- Sequencing agents reduced friction in navigation.
- Nudging agents reduced procrastination.
- Reinforcement agents provided rapid feedback.
- Friction reduction agents reduced switching costs.

Trust Calibration and Absence of Algorithm Aversion

The absence of completion, collapse, or performance degradation suggests that algorithm aversion effects were not dominant in the observed behavior (Dietvorst et al., 2015).

Several structural factors likely contributed to the calibration of trust, which could have been eroded through algorithm aversion effects:

- The micro-agents did not have evaluative control.
- Managers retained control.
- The system provided nudges, not prescriptive recommendations.

These characteristics are consistent with the set of human-centered AI system design principles, which emphasize transparency and controllability (Amershi et al., 2019). By constraining algorithmic control within a human-in-the-loop framework, this deployment avoids conditions likely to lead to algorithm aversion. Trust calibration was not empirically measured but is based on inference.

Human Oversight as Governance Infrastructure

Human oversight served as a contextual stabilizer rather than a statistically controlled moderator of algorithmic effects. Managers retained control for monitoring progress, interpreting analytics, and addressing issues throughout both phases of the deployment.

Although the intensity of oversight was not empirically quantified, the co-existence of moderate efficiency gains (d = -0.52) and stable assessment performance suggests that governance likely mitigated superficial acceleration or metric gaming.

This finding supports responsible AI governance frameworks (Papagiannidis et al., 2025) and extends the logic of augmentation-based integration (Raisch & Krakowski, 2021). AI-enhanced training systems likely have the greatest sustainability when embedded in responsible managerial structures.

Future studies should measure intensity of oversight using intervention frequency or audit trail measures.

Alternative Explanations

There are several alternative explanations for the results which must be considered:

- Learning Curve Effects: The improvements may be related to the learning curve rather than the micro-agents alone. Yet, the sustained improvements in assessment scores suggest the time-on-task improvements were not simply related to navigation.
- Organizational Expansion: The expansion of outlets from 35 to 50 has introduced heterogeneity within the workforce. In such cases of expansion, it would be expected that volatility would increase. The absence of this volatility suggests resiliency but does not allow for inference.
- Novelty Effects: The pilot period would be expected to be characterized by novelty effects. The sustained nature of the improvements beyond the pilot period suggests they are not simply related to novelty.

These considerations serve to further support the inference, which was anticipated given the nature of the research approach.

Theoretical Contributions

This research has made several important contributions to the relevant theories:

- This research has added to the technology acceptance body of knowledge by demonstrating the effect of technology deployment maturity on normalization and efficiency outcomes (Davis, 1989; Venkatesh & Davis, 2000).

- This research has added to the training body of knowledge by providing objective measures of behavioral outcomes rather than perceptual self-reporting (Baldwin & Ford, 1988; Podsakoff et al., 2003).
- By considering AI systems as being comprised of micro-agents within the structure of governance, this research has added to the human-AI body of knowledge by transcending the dichotomy of automation and autonomy. The research suggests that constrained micro-agents may be beneficial for sustainable patterns of performance.

Practical Implications

In terms of practical implications for practitioners, it should be recognized that normalization will occur and should be encouraged rather than fought. Perfect pilot completion rates are not expected to be sustained during expansion.

More critically, the moderate efficiency effect (d = -0.52) suggests that AI-driven micro-agents can significantly decrease time costs per unit of learning, thereby supporting scalability. However, efficiency should be balanced with the need for continuous governance especially when ensuring the integrity of the learning process. The more systems speed up processes, the more they require more and not less oversight.

Limitations and Future Research

There are limitations to this research, and these limitations provide a framework for future research.

Contextual and Generalizability Limitations

One of the first limitations is that, despite its focus on a rapidly expanding Italian casual dining restaurant chain, the research is specific to a particular context. In addition, the fact that the organization expanded from 35 outlets to 50 outlets during the course of the research is a contextual constraint. The impact of expansion can influence learning outcomes.

Absence of Experimental Control and Limitations in Causal Inference

Secondly, the current study employs an observational rather than an experimental or quasi-experimental research framework. Although this improves ecological validity, causal attribution is limited. The observed improvements in learning efficiency and assessment outcomes are described as statistically significant rather than causal associations of AI-based micro-agents with learning efficiency and assessment outcomes.

Potential confounding variables:

- Changes in workforce composition
- Variability in tenure
- Variability in management enforcement

- Growth of organizational units

Future studies could use quasi-experimental designs to isolate causal effects.

Hawthorne and Novelty Effects

Thirdly, the 2024 pilot phase may have experienced Hawthorne or novelty effect bias, where employees participating in the doctoral study may have recorded higher engagement due to raised visibility and possibly evaluation of their behavior or novelty of the digital tool.

Though sustained improvements in efficiency observed during the 2025 phase minimize novelty effect bias, as observed in the current study, it is not possible to rule out novelty effect bias entirely as observed during the 2024 phase.

Future studies could use baseline or delayed treatment groups to help disentangle novelty effect bias from behavioral adaptation.

Platform Dependency and System-Specific Effects

Fourth, it is important to note that this study was based entirely on the Safety Culture AI-based training platform. While this platform represents a constrained, workflow-based AI micro-agent, it is possible that platform-specific effects may have contributed to observed effects. Future research should consider comparing multiple AI-based training systems in order to better understand the effects of micro-agent design, transparency, configuration, and autonomy levels on adoption trajectories and efficiency gains.

Temporal Scope and System Maturation

Although the current research covers pilot and scaled phases over two years, it does not cover long-term post-adoption system maturation beyond the scaled phase. Information systems research has emphasized the importance of post-adoption and routinization behaviors, which take place over long periods and can include effects such as habituation, oversight recalibration, efficiency plateauing, and even engagement decay. Long-term longitudinal research would be useful for studying:

• Habituation

• Oversight recalibration

• Efficiency plateauing

• Engagement decay

These are interesting areas for future research on post-adoption information systems maturation.

Future Research Agenda

Based on the limitations of the current research, future research should be conducted in the following four directions:

1. Experimental Micro-Agent Manipulation

Randomly activate and deactivate particular micro-agent features, for example, nudging, sequencing, and reinforcement.

2. Governance Intensity Modeling

Quantify oversight and model it in the regression model framework for moderation effect testing.

3. Contextual Growth Moderation

Model organizational expansion variables in the regression model framework for moderation effect testing.

4. Cross-Platform Comparative Research

Explore the effect of differences in AI-enabled training architectures on longitudinal adoption and efficiency patterns.

Conclusion

In this paper, the performance of AI-powered training micro-agents as they transition from the pilot stage to the scaled stage within a service-intensive environment characterized by front-line employees is examined. By employing objective and trace-based behavioral data, it analyzed adoption patterns, learning outcomes, and time efficiency between the two stages of micro-agent deployment.

The results produced three primary findings:

First, the completion rate normalizes under scaled deployment. While the pilot stage was able to attain a 100% completion rate ceiling, the scaled stage attains a still-high but more realistic 86.82% completion rate. This does not suggest disengagement but rather more realistic adoption patterns as the operations transition from the pilot stage to the scaled stage. What should be emphasized here is the fact that engagement still remains robust even under scaled deployment pressure.

Second, learning outcomes improve under scaled deployment. The average assessment scores are substantially higher under the scaled stage. This shows that repeated exposure and integration with operations improve learning outcomes. This finding counters the conventional assumption that learning outcomes deteriorate under scaled deployment.

Third, learning outcomes are more efficient under scaled deployment. The time to complete the assessment reduces under the scaled stage. This shows that employees are more familiar with the micro-agents and the content they are promoting. The

combined improvement of learning outcomes and efficiency shows us that the micro-agents are maturing and not deteriorating under scaled deployment.

Theoretically, this research makes a contribution to information systems and organizational learning research by bringing together training transfer theory, the technology adoption model, and human-in-the-loop AI governance within a longitudinal and real-world setting. The research extends existing research in this area by illustrating that it is long-term exposure and integration, rather than initial adoption, that influence long-term effectiveness of AI-based training systems. The use of exposure-adjusted active employees represents a specific methodological approach that can inform analysis of adoption in a high-turnover setting.

Practically, this research illustrates that success of an AI-based training system cannot be solely defined by initial pilot metrics. For example, initial conditions of an AI-based system may artificially inflate completion rates due to increased oversight and novelty effects. Rather, success of an AI-based system should be defined by normalized engagement rates, stable or improving metrics, and increased efficiencies over time.

While this research is specific in its setting and does not use a randomized control group, this research offers a longitudinal and behavioral-based approach that examines the effectiveness of an AI-based training system within a real-world setting and under actual scaling conditions. Future research should seek to extend this research by conducting a multi-industry analysis and utilizing a hierarchical approach to validate this research in a different setting.

In conclusion, this research illustrates that micro-agents of an AI-based training system do not degrade in effectiveness as scale increases. Rather, this study shows that an AI-based training system stabilizes, matures, and improves inefficiencies as integration increases within an organizational setting.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., … Horvitz, E. (2019). *Guidelines for Human-AI Interaction*. CHI 2019. https://doi.org/10.1145/3290605.3300233

Aradhyula, G. (2024). *Human-in-the-loop machine learning systems. World Journal of Advanced Engineering Technology and Sciences*, 11(1), 514–521. https://doi.org/10.30574/wjaets.2024.11.1.0012

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*(1), 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics, 13*(1), 333–372. https://doi.org/10.1257/mac.20180386

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dote Pardo, J. S. (2025). Artificial intelligence in organizational learning: A systematic review of applications, opportunities, and challenges. *Development and Learning in Organizations: An International Journal*. https://doi.org/10.1108/DLO-06-2025-0228

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2019). Cognitive challenges in human–AI collaboration (working paper). SSRN. https://doi.org/10.2139/ssrn.3368813

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human–AI collaboration. *European Journal of Information Systems*. https://doi.org/10.1080/0960085X.2025.2475962

Hosen, S., Hamzah, S. R., Ismail, I. A., Alias, S. N., Abd Aziz, M. F., & Rahman, M. M. (2024). Training & development, career development, and organizational commitment as the predictor of work performance. *Heliyon, 10*(1), e23903. https://doi.org/10.1016/j.heliyon.2023.e23903

Madanchian, M., Taherdoost, H., & Mohamed, N. (2023). AI-based human resource management tools and techniques: A systematic literature review. *Procedia Computer Science, 229*, 367–377. https://doi.org/10.1016/j.procs.2023.12.039

Muehlemann, S. (2025). *Artificial intelligence adoption and workplace training. Journal of Economic Behavior & Organization*, 238, 107206. https://doi.org/10.1016/j.jebo.2025.107206

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). *Responsible artificial intelligence governance: A review and research framework*. Journal of Strategic Information Systems, 34, Article 101885. https://doi.org/10.1016/j.jsis.2024.101885

Pappas, I. O., Giannakos, M. N., & Sampson, D. G. (2019). Fuzzy set analysis for learning systems: The role of complex concepts and human factors. *Computers in Human Behavior, 92*, 646–659. https://doi.org/10.1016/j.chb.2018.03.032

Passmore, J., & Daly, J. (2026). Are AI coaching agents learning friend or foe?: A qualitative study of learning and development leaders' perceptions of AI coaching agents using Unified Theory of Acceptance and Use of Technology 2. *International Journal of Training and Development*. https://doi.org/10.1111/ijtd.70028

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model. *Management Science, 46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412