

## Strategic Human Oversight Frameworks for AI-Enabled Training Microagents: Evidence from a Longitudinal Adoption Study

Smrite Goudhaman

Faculty of Business Analytics, Golden Gate University, San Francisco, California, United States.

[smritegoudhaman@gmail.com](mailto:smritegoudhaman@gmail.com)

### RESEARCH ARTICLE

### Open Access

#### ARTICLE INFORMATION

Received: January 19, 2026  
Accepted: March 25, 2026  
Published Online: April 15, 2026

*Keywords:*

AI enabled training, Micro agents, AI governance, Longitudinal study, Technology adoption

#### ABSTRACT

**Purpose:** The purpose of this study is to examine how AI-enabled training outcomes evolve as an intervention transition from a supervised doctoral pilot to a scaled, longitudinal organizational deployment. The study focuses on learning adoption and learning efficiency in a frontline hospitality context, while explicitly examining the role of AI micro-agents operating within a human-in-the-loop governance framework.

**Methods:** The study adopts a longitudinal cohort extension design, building on a doctoral pilot conducted with 100 frontline employees during 2024 and extending into a scaled operational deployment during 2025. Objective learning-platform trace data from an AI-enabled training system were analyzed across two deployment phases. Learning adoption was measured using exposure-adjusted completion rates, while learning efficiency was assessed using assessment performance and time-on-task metrics. The analysis controls for workforce churn characteristic of frontline service environments.

**Findings:** Results show that completion rates normalized from 100% in the pilot phase to 86.82% under real-world scale, reflecting operational normalization rather than reduced effectiveness. Importantly, learning quality and efficiency improved over time: mean assessment scores increased, while average time-on-task declined significantly. These findings indicate faster mastery and deeper learning as AI-enabled training matured, rather than superficial compliance.

**Implications:** The findings demonstrate that AI-enabled training systems can sustain adoption and improve learning efficiency at scale when designed with constrained agency and supported by human oversight. For organizations in high-churn frontline environments, the results emphasize the importance of evaluating training effectiveness beyond pilot completion metrics and focusing on longitudinal learning quality, efficiency, and governance structures.

**Originality:** This study provides rare longitudinal, post-dissertation evidence on AI-enabled training effectiveness, directly linking a doctoral pilot to scaled organizational deployment. It advances technology management and digital learning research by introducing a churn-aware evaluation framework and empirically demonstrating how AI micro-agents, operating within human-in-the-loop governance, shape sustainable learning outcomes beyond pilot conditions.



DOI: [10.15415/jtmge/2025.162007](https://doi.org/10.15415/jtmge/2025.162007)

## 1. Introduction

The integration of Artificial Intelligence (AI) is changing the organizational learning architecture in various industries, including training environments, especially in industries where training is operationally significant at the front lines. For instance, there is increased use of AI systems to automate training programs, tailor training programs to individual learners' needs, and analyze learners' engagement patterns in real time. AI systems are no longer limited to digital training programs; rather, they are now AI micro-agent systems, which have the power to influence behavioral outcomes.

Recent studies have shown that the adoption of AI systems is associated with training investment in the workplace, skills development in the workforce, and the evolution of the learning architecture in organizations (Muehleman, 2025). Thus, the adoption of AI systems in training environments is not limited to the current period; rather, it is associated with the evolution of the structural mechanisms that influence the training learning architecture.

Although there is an increasing trend towards investing in AI systems, few studies have been conducted to investigate the impact of the longitudinal integration of AI systems in

training environments on training adoption, assessment, and completion efficiency. Additionally, there is limited literature examining the impact of the use of humans in AI systems.

This study aims to bridge the identified gap in the literature by examining the longitudinal impact of AI system micro-agent deployment maturity in training environments on training adoption, assessment, and completion efficiency in the hospitality industry.

### *1.1. The Effectiveness of Training beyond the Pilot*

Classic studies on the effectiveness of training programs have indicated that the learning process and the implications for the entire organization are critical aspects to consider in training effectiveness (Baldwin & Ford, 1988). The pilot program, as a critical aspect of training effectiveness, also raises important considerations concerning the entire environment and the implications for the organization.

This is especially critical in a service environment in which the employee might not only terminate the training process but also might not start the training process in the first place or might not make the entire learning process a priority due to the environment and the need to perform the job.

### *1.2. The Effectiveness of Technology beyond the Pilot*

Classic studies on the effectiveness of technology and its acceptance by the entire organization and its employees have indicated that the entire learning process and the environment are critical aspects to consider in technology effectiveness and acceptance by employees and the organization (Davis, 1989; Venkatesh & Davis, 2000; Venkatesh *et al.*, 2003).

Even though these studies provide critical insights concerning the initial stages of technology adoption and acceptance by employees and the organization, there is limited knowledge concerning the long-term evolution and implications over extended periods. The initial stages of training might not guarantee the long-term process in a training environment; thus, a longitudinal study is necessary to separate the two factors: adoption and the entire learning process.

### *1.3. AI-Enabled Training, Micro-Agents, and Human Oversight*

With developments in AI-enabled human resource management and learning management, there is a need to

design micro-level interventions to guide user behavior. AI-enabled micro-agents have also been designed to assist users in completing tasks and learning by providing reminders. Such micro-level interventions have been designed to support learning and performance by reducing cognitive and operational barriers.

Although such micro-level interventions have been designed to support learning and performance by reducing cognitive and operational barriers, their implications within governed training systems remain underexplored.

In the context of training systems, such issues become particularly relevant. The implications of training outcomes on service quality, security, and effectiveness are considerable. As such, it becomes essential to evaluate AI-based training systems from a governance perspective, i.e., the interaction between AI-based micro-agents and human-in-the-loop systems for training integrity.

### *1.4. From Doctoral Pilot to Longitudinal Evidence*

The present research extends from previous doctoral research into 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.

Subsequent to completing 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 an environment beyond initial conditions.

### *1.5. Rethinking Learning Outcomes in High-Churn Environments*

The effectiveness of training in frontline environments faces challenges owing to the high churn rate of employees. This is because employees may leave an organization before completing their scheduled training. Additionally, employees may join an organization when training is already underway. A learning system captures all training interactions, including those that are incomplete. This may affect results if exposure is not taken into account when evaluating outcomes.

This paper, therefore, proposes an alternative approach by considering only active employees, i.e., employees who have been active in learning within specific periods. This provides a contextual interpretation of training results, i.e., an emphasis on the quality and efficiency of training,

including test results and time spent on tasks. This provides a more objective measure of training results compared to subjective self-reporting, which faces a variety of method biases (Podsakoff *et al.*, 2003).

### 1.6. 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 learning adoption and learning efficiency as the program transitions from the pilot phase to the scaled deployment phase, where AI-enabled micro-agents 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 and learning adoption and learning efficiency as the dependent variables, allowing the study to explore the post-pilot performance of the system.

This paper makes three contributions to the digital information systems field. 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 interpreting learning adoption and learning efficiency outcomes. Third, the study extends the understanding of AI micro-agents, human collaboration, and learning performance beyond the pilot phase.

### 1.7. Structure of the Paper

This paper is structured as follows. Section 2 reviews the literature on training effectiveness, technology adoption, AI-enabled learning, and human–AI collaboration. Section 3 presents the research model and research hypotheses. Section 4 presents the research design, while Section 5 presents the research findings. Section 6 presents the discussion, and Section 7 concludes the paper.

## 2. Literature Review

### 2.1. 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 at a single point in time. Baldwin & Ford's foundational research on transfer of training has recognized that the effectiveness of learning outcomes over time is dependent on how learning is reinforced, applied, and supported over time. Though the effectiveness of learning outcomes may be high in the short term, effectiveness tends to diminish over time as organizational conditions change.

This is particularly the case in frontline settings, where employees are under time pressure, exhibit role variability, and experience high turnover. In such settings, it is critical that training systems not only provide learning outcomes but also sustain employee engagement and learning behavior as employees join and leave the organization. Hence, longitudinal evaluation is critical in evaluating the effectiveness of learning outcomes, rather than single-point evaluation.

The mechanisms provided by artificial intelligence systems hold promise in bridging the persistence gap through adaptive reinforcement, feedback, and continuous monitoring of engagement. Personalization engines, feedback, and adaptive content systems, which are the core mechanisms in the application of artificial intelligence in organizational learning, were highlighted in a recent systematic review on the application of artificial intelligence in organizational learning (Dote Pardo, 2025). These mechanisms promote continuous engagement rather than a singular exposure.

The lagged realization of performance benefits following the adoption of technology is a widely accepted aspect in economic theories. Brynjolfsson, Rock, and Syverson (2021) describe this aspect in the following words: "This lagged realization of performance benefits is often referred to as the productivity J-curve. An organization must make investments in adjustment costs before it can realize the performance benefits associated with general-purpose technologies." This is applicable to the adoption of AI training systems as well; it is possible that the initial pilots may not have a significant impact on the adoption process, while such impacts are realized at a later stage.

However, despite considerable research on the transfer of learning outcomes, few studies have longitudinally evaluated learning outcomes in pilot and field studies. Most studies have evaluated the effectiveness of learning outcomes either shortly after the completion of the learning program or within a very narrow time window.

## 2.2. 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 in 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. This differentiation is particularly important when considering training programs where initial compliance with a program may be a result of structural supervision rather than intrinsic engagement. Thus, a longitudinal approach must be taken to understand the differentiation between initial compliance and eventual patterns of adoption. Without such differentiation, an organization may overestimate the efficacy of a training program based on initial patterns of engagement.

Research in technology acceptance has long demonstrated that user engagement is a function of perceived usefulness, effort expectancy, social influence, and facilitating conditions. The Unified Theory of Acceptance and Use of Technology (UTAUT) and its subsequent extensions have provided a robust model for understanding behavioral intention and usage patterns in organizational contexts (Venkatesh *et al.*, 2012). These factors are of particular importance in an AI-based training environment, where adoption is a function of technological capability and system legitimacy within existing organizational workflows. Therefore, maturity could have a potential impact not just on levels of exposure but also on the behavioral dynamics of acceptance.

Recent qualitative research examining the effectiveness of AI coaching agents using the Unified Theory of Acceptance and Use of Technology 2 model found that performance expectancy, effort expectancy, and trust perceptions were significant factors in the willingness of learning leaders to use AI systems (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.

## 2.3. AI-Enabled Training and Micro-Level Interventions

Significant advancements have been made in AI-enabled human resource and learning systems. These advancements have led to the development of micro-level interventions that can support user behavior. For instance, micro-level interventions can be used as reminders and cues that can be directly incorporated into workflows. These micro-level interventions are designed to support user behavior. They are not intended to replace decision-making but rather to reduce the level of friction and encourage users to engage with a desired course of action.

AI-based micro-agents work at the micro-intervention level to affect user behavior through nudging, adaptive sequencing, and reinforcement prompts incorporated into the training process. Unlike static training programs, these systems are dynamic in nature, as they respond to signals of user involvement.

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

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 structure learning and determine the order and timing of tasks.
3. Reinforcement Agents: Agents that reinforce learning through prompts and feedback.
4. Friction-Reduction Agents: System-level changes that reduce cognitive or operational hurdles (e.g., autosaving, micro-quizzes, easy navigation).

These categories are associated with behavioral reinforcement theory and best practices in digital workflow design. This provides insight into how AI micro-agents operate at a task level rather than a strategic level. Conceptualizing micro-agents in a structured manner enables more precise analysis when examining their long-term effects on employees. In a learning environment, such micro-agents have the potential to influence employees' use of training materials, which may increase the success rate of training program completion. However, such success is not solely based on the algorithms

used to program micro-agents but also on employees' perceptions of them.

Algorithm aversion studies suggest that employees are more likely to be dissatisfied with, and even terminate, algorithm-based interactions if they observe errors or inconsistencies (Dietvorst *et al.*, 2015). In organizational contexts, there are often conflicting factors between automation and augmentation. Rather than viewing artificial intelligence as a replacement for managerial supervision, contemporary theory suggests understanding AI as an augmentation of human decision-making processes with supervisory authority. This has been discussed by Raisch and Krakowski (2021), where the need to address the automation–augmentation paradox was emphasized as essential for balancing human interpretability with efficiency. In training environments, micro-agents are therefore better conceptualized not as evaluators but as augmentation tools that assist supervisors. This perspective aligns with current research and governance-oriented approaches to AI.

#### 2.4. Human–AI Collaboration and the Role of Oversight

An emerging research paradigm supports the notion that AI micro-agents are best implemented as tools that complement human judgment, as opposed to tools that are implemented independently. Research on human–AI collaboration highlights the cognitive challenges that emerge when human and AI micro-agents are not appropriately defined (Fügener *et al.*, 2019). More recent research supports the notion that AI micro-agents are best implemented as tools that complement human judgment, in which human strengths are augmented by AI micro-agents, while humans provide the necessary oversight (Hemmer *et al.*, 2025).

Guidelines for human–AI collaboration emphasize the need for human awareness, the ability to intervene, and the presence of accountability systems in human–AI interactions (Amershi *et al.*, 2019).

With regard to human oversight in AI-enabled training systems, such oversight includes monitoring learner progress, handling exceptions, and reinforcing training objectives while remaining consistent with organizational goals. In the absence of human oversight, interventions by micro-agents are likely to be ineffective or even counterproductive.

#### 2.5. Integrative Synthesis: Training, TAM, and Human–AI Systems

Research on training effectiveness, technology adoption, and human–AI collaboration has largely evolved in parallel. However, to gain a comprehensive and integrative understanding of AI-enabled training systems, these literatures must be synthesized.

The literature on training effectiveness has highlighted reinforcement and transfer mechanisms (Baldwin & Ford, 1988). Technology acceptance literature has highlighted the determinants of initial adoption and long-term usage patterns (Davis, 1989; Venkatesh & Davis, 2000; Venkatesh, Morris, & Davis, 2003). Finally, literature on human–AI collaboration has highlighted the governance structures required to facilitate such interactions (Fügener *et al.*, 2019; Hemmer *et al.*, 2025; Amershi *et al.*, 2019).

Training systems exist at the intersection of all three disciplines. Micro-agents serve as the mechanism through which reinforcement is carried out at scale. Technology acceptance factors are crucial in determining the effectiveness of reinforcement interventions. Finally, human oversight plays a critical role in determining the extent to which algorithmic reinforcement is trusted and aligned with organizational goals.

The effectiveness of longitudinal AI-enabled training systems is enabled through three mechanisms: behavioral reinforcement (training transfer theory), technology adoption and usage (TAM/UTAUT), and governance and oversight complementarity (human–AI systems theory).

#### 2.6. Measuring Learning Outcomes: Beyond Completion Rates

The majority of training studies are based on self-reported data and single-item measures such as completion rates. However, such studies are vulnerable to common method bias and may not accurately measure learning outcomes (Podsakoff *et al.*, 2003). In contrast, data from learning platforms are more accessible and provide more accurate information for determining training effectiveness.

Assessment scores are crucial in determining learning effectiveness, while time spent on tasks is crucial in determining learning efficiency. A decline in time spent on tasks alongside constant or improving assessment scores is indicative of increased learning efficiency rather than a lack of engagement (Pappas *et al.*, 2019). Such data are particularly important in longitudinal studies, as they may indicate the maturation of the learning system and increasing user familiarity.

In the context of the frontline workforce, where there is significant churn, it is also important to consider the exposure effect. Measuring learning effectiveness for employees who are actively engaged reduces bias arising from partial engagement.

#### 2.7. Positioning the Present Study

Although prior research has highlighted the potential of AI-supported learning systems, there is a lack of research examining the effects of deployment maturity on the

performance of such systems over extended periods. Most studies have focused on the initial stages of deployment or on perceptions of system effectiveness.

To bridge this research gap, this study aims to determine whether the deployment of AI-supported micro-agents in a training environment is more effective compared to a supervised pilot phase.

## 2.8. Research Questions

- RQ1: Does the deployment of AI-supported micro-agents increase course completion rates compared to the supervised pilot phase?
- RQ2: Does the deployment of AI-supported micro-agents increase mean assessment performance compared to the supervised pilot phase?
- RQ3: Does the deployment of AI-supported micro-agents decrease mean time-on-task compared to the supervised pilot phase?

## 2.9. Hypotheses

Based on 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.

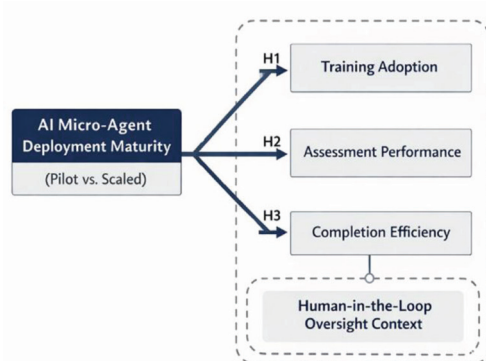


Figure 1: Conceptual Research Model

## 3. Research Model and Research Questions

### 3.1. 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 systems and training literature, 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 centres 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.

### 3.2. 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 in 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 setting (pilot) to a naturalistic and widespread setting (scaled).

This process is in line with earlier studies on post-adoption system maturation and usage transitions (Venkatesh & Davis, 2000). It allows for a statistical evaluation of learning outcomes under different deployment conditions, with the underlying configuration of the AI system remaining constant.

### 3.3. 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 learning adoption and learning efficiency, addressing criticisms of evaluation methods that focus on a single metric.

#### 3.3.1. Learning Adoption – Completion Rate

Learning adoption is measured by completion rate, defined as:

$$\text{Completion Rate} = \frac{\text{Completed Records}}{\text{Assigned Records}}$$

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

Completion is interpreted contextually, with pilot completion reflecting structured supervision and scaled completion reflecting 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 that completion is a fair reflection of actual participation, rather than being confounded by attrition effects.

#### 3.3.2. Learning Quality – Assessment Performance

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

$$\text{Mean Time-on-Task} = \frac{\text{Sum of Time-on-Task}}{\text{Completed Records}}$$

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.

#### 3.3.3. Learning Efficiency – Time-on-Task

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

$$\text{Mean Time-on-Task} = \frac{\text{Sum of Time-on-Task}}{\text{Completed Records}}$$

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

Time-on-task is interpreted alongside assessment performance. If time-on-task decreases while assessment

performance is stable or increasing, this suggests that learning is becoming more efficient rather than reflecting disengagement (Pappas *et al.*, 2019).

Together, assessment performance and time-on-task provide a balanced and behaviorally grounded view of learning effectiveness.

### 3.4. 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 (Fügenger *et al.*, 2019; Hemmer *et al.*, 2025).

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. This clarification ensures conceptual precision and removes potential validity issues related to measurement of the construct.

### 3.5. 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: Mean assessment performance varies between the supervised pilot phase and the scaled operational phase.
- H3: Mean time-on-task varies between the supervised pilot phase and the scaled operational phase.

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

### 3.6 Research Model Overview

#### 3.6.1. Conceptual Research Model

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

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

There are three research hypotheses for this model:

- 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.

The research model assumes these relationships within a human-in-the-loop context of AI micro-agent training deployment. This context conceptualizes AI micro-agent training as technological augmentation rather than a replacement for human managerial judgment. It involves the incorporation of ethical considerations and performance monitoring for AI training and its outcomes. Thus, Figure 1 illustrates the directional relationships between AI micro-agent training and its outcomes, as postulated by technology adoption and AI governance theories.

#### 3.6.2. Research Methodology Flow

Figure 2 shows the research methodology flow of the research process for testing the proposed model. The research process consists of six phases:

1. **Organizational Context and Research Objective** – Identification of training challenges in the frontline hospitality organization and development of research objectives.
2. **Pilot Phase Implementation (Months 1–3)** – Initial implementation of AI-based micro-agents with the pilot group, establishing foundational metrics.

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 variables:
  - **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 statistical methods:
  - Two-proportion z-test for differences in adoption rates
  - Welch's t-test for differences in 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 findings to the theoretical foundations of the micro-agent model, AI governance, and broader implications for technology management in training systems.

Figures 1 and 2 distinguish the theoretical and operational aspects of the study. Figure 1 describes the theoretical model defining relationships between variables, whereas Figure 2 outlines the operational steps taken to test those relationships.

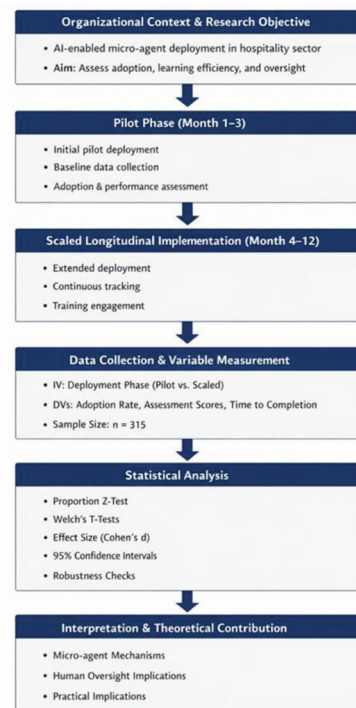


Figure 2: Research Methodology Flowchart

## 4. Research Design and Methodology

### 4.1. 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.

In this study, AI-enabled training is conceptualized not as a simple instructional intervention, but rather as a set of constrained, task-specific micro-agents that are embedded into the organizational workflow, monitoring learner response patterns while implementing pre-specified interventions, while at the same time working within boundaries that are explicitly defined, supervised, and controlled by human managers. The longitudinal design is particularly pertinent to this study, given the expected evolution of the behavioral impact of the micro-agents as the intervention becomes more embedded into the organizational workflow.

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.

### 4.2. 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 environment. In the context of the pilot study conducted in 2024, Toscano was planning to expand from 35 outlets to 50 outlets, creating high operational pressure to scale up frontline training while maintaining consistency in service quality, food safety, and customer experiences.

This expansion context is significant from an analytical perspective, as rapid outlet expansion usually accentuates training issues common within the hospitality industry, including the quality of onboarding, skills gaps between staff and specific outlets, and the number of first-time managers. When Toscano was set to expand its outlets, the organization recognized the importance of training its frontline staff, specifically the need for training consistency, which was considered a critical risk factor that could not be effectively managed using manual methods.

### 4.3. 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 frontline staff from selected outlets.

As a digital training platform, SafetyCulture was used by Toscano as an AI-enabled training environment, where the organization was able to use structured learning modules, automated task assignment, and training analytics to track the performance and behavior of frontline staff. While the platform was able to generate recommendations, it did not exercise decision-making authority, and all decisions regarding the interpretation of staff behavior and performance were made by human managers.

Following the completion of the pilot phase, Toscano expanded the system in 2025 in parallel with its outlet expansion, extending the AI-based training system across its employee population. Notably, the content of the training and the configuration of micro-agents were kept unchanged to facilitate longitudinal comparison. The scaled phase represented a shift from experimental use to organizational use of the system, thereby creating an environment where its use would naturally evolve from being perceived as new to becoming familiar.

### 4.4. 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.

The initial number of employees enrolled was approximately 100. Employees considered active were those who demonstrated some level of engagement, defined as completing at least one assignment record.

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

Inferential analysis was conducted on the full pilot exposure dataset, i.e., 886 records.

### 4.5. Measures

Three measures were defined based on the data obtained from the SafetyCulture 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 rather than a statistically controlled moderator.

#### 4.6. Analytical Approach (Upgraded)

Descriptive analyses (mean, standard deviation, variance, minimum, maximum) were conducted separately for each phase. Inferential tests were conducted to ensure analytical rigor:

- **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 conducted as a robustness test:  
$$\text{Outcome}_i = \beta_0 + \beta_1 \text{DeploymentMaturity}_i + \varepsilon_i$$
  
Deployment Maturity = 0 (pilot), 1 (scaled)  
Results were consistent with t-test findings.

#### 4.7. Control Variables and Organizational Context

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

- Core training content is constant.
- Micro-agent configuration is constant.
- Expansion throughout 2025 is gradual rather than abrupt.

Future research should incorporate structured employee-level controls such as tenure, role type, and outlet age.

#### 4.8. Survivorship and Cohort Considerations

The data does not track a panel of identical individuals across both phases, due to workforce changes arising from organizational expansion and attrition.

As a result:

- Longitudinal effects are observed at the organizational level.
- Results should not be interpreted as causal processes at the individual level.
- Exposure-adjusted definitions of active users reduce, but do not eliminate, the effects of churn.

#### 4.9. Causal Inference Limitations

Given the non-experimental design:

- There is no randomization.
- There is no counterfactual comparison group.
- Organizational expansion coincides with deployment scaling.

As a result, findings should be interpreted as statistically significant longitudinal associations, rather than causal effects of AI-enabled training deployment.

### 5. 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. This approach provides a better representation of systemic behavioral patterns rather than individual-level inference.

#### 5.1. 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, training volume and workforce engagement increased substantially with outlet growth.

## 5.2. 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 vs. 2024 pilot phase): OR = 0.16

The odds ratio indicates that the likelihood of course completion under scaled deployment is significantly lower than that observed under the tightly supervised 2024 pilot phase. However, the substantial increase in training volume and workforce engagement indicates normalization under scaled deployment rather than disengagement.

## 5.3. 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  $d = 0.14$

This represents a small but statistically significant improvement in assessment performance for scaled deployment.

## 5.4. Learning Efficiency: Time-on-Task

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

### 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  $d = -0.52$

This represents a moderate improvement in efficiency, where learners were able to complete modules in significantly less time with no decrease in assessment performance.

## 5.5. Robustness Checks

To test robustness, several 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 outlets where role-level identifiers were available, similar effects were observed.

## 5.6. Alternative Explanations

- **Learning Curve / Familiarity Effect:** The efficiency gain may be partly attributed to increased familiarity with the platform rather than solely the effectiveness of micro-agents. Nevertheless, increased assessment performance alongside decreased time-on-task indicates mastery effects.
- **Organizational Expansion Effects:** Toscano expanded its outlets from 35 to 50 during the study period. This expansion may have introduced variability due to new users. Nevertheless, consistent efficiency gains across conditions suggest positive adaptation.
- **Reduced Novelty / Hawthorne Effects:** Novelty effects in the pilot condition may have contributed to increased learning behavior. Nonetheless, sustained high completion rates (86.82%) and improved performance indicate that learning behavior persisted beyond novelty effects.

The interpretations presented are consistent with a non-experimental design and are conceptualized as longitudinal effects rather than causal relationships.

## 5.7. Summary of Longitudinal Outcomes

In summary, the results of the analysis indicate a consistent longitudinal pattern in the data collected from the AI-enabled learning system as it transitioned from pilot to large-scale deployment:

- Completion rates normalized in large-scale deployment while remaining high.
- Learning effectiveness increased, as evidenced by statistically significant improvements in assessment performance ( $d = 0.14$ ).

- Learning efficiency improved in large-scale deployment, as evidenced by moderate reductions in time-on-task ( $d = -0.52$ ).

The results indicate that large-scale deployment of the learning system is statistically associated with sustained adoption and improved learning efficiency.

## 6. 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.

### 6.1. From Pilot Optimality to Scaled Normalization

A significant finding is the statistically significant reduction in completion rates from 100% during the supervised pilot to 86.82% during the scaled rollout ( $z = 11.52$ ,  $p < .001$ ). Although cross-sectional analysis suggests a decline during rollout, the longitudinal analysis indicates normalization.

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

From the perspective of technology acceptance theory, the fact that more than 85% of the population completed the training at such expanded volume indicates sustained engagement. It is also noteworthy that the conditions of the pilot, although optimal, are not representative of real-world environments, where conditions are less controlled. Technology acceptance theory suggests that, in real-world settings, adoption rates typically normalize.

This finding also supports the “J-curve” framework on productivity dynamics, which posits that organizational performance following technology integration often follows a normal adjustment pattern.

### 6.2. Learning Efficiency as a Longitudinal Association

The most theoretically interesting outcome is that both assessment scores and efficiency improved. The improvement

in assessment scores was small ( $d = 0.14$ ) but statistically significant, whereas the decrease in mean time-on-task showed a moderate effect size ( $d = -0.52$ ).

The distinction in effect size is noteworthy. While the improvement in assessment scores is small, its statistical significance indicates consistency. In contrast, the moderate reduction in time-on-task reflects a more meaningful behavioral outcome. This supports the argument that deployment maturity is more strongly associated with efficiency optimization than with dramatic gains in knowledge acquisition.

From a transfer-of-training perspective (Baldwin & Ford, 1988), this pattern aligns more closely with procedural fluency than with score inflation.

The observed effect sizes suggest that micro-agents operated primarily at the workflow level rather than at the content level. This aligns with human–AI complementarity frameworks (Fügener *et al.*, 2019; Hemmer *et al.*, 2025). The moderate effect size further indicates that the outcome is substantively meaningful, not merely statistically significant.

### 6.3. Micro-Agent Mechanisms and Augmentation Logic

Several micro-agent mechanisms may have contributed to these outcomes:

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

These mechanisms operate at the level of behavioral architecture and do not directly affect knowledge content. The findings are consistent with augmentation theory and do not support automation theory (Raisch & Krakowski, 2021). The micro-agents did not substitute for evaluative authority.

The absence of performance inflation alongside efficiency gains further suggests that acceleration effects did not compromise learning integrity.

### 6.4. Trust Calibration and Absence of Algorithm Aversion

The absence of completion collapses 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 trust calibration, which might otherwise have been undermined by algorithm aversion:

- The micro-agents did not have evaluative control.
- Managers retained decision-making authority.
- The system provided nudges rather than prescriptive recommendations.

These characteristics are consistent with human-centered AI 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 induce algorithm aversion. Trust calibration was not empirically measured and is therefore inferred.

### 6.5. Human Oversight as Governance Infrastructure

Human oversight served as a contextual stabilizer rather than a statistically controlled moderator of algorithmic effects. Managers retained responsibility for monitoring progress, interpreting analytics, and addressing issues throughout both phases of 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 are likely to be most sustainable when embedded within responsible managerial structures.

Future studies should measure oversight intensity using indicators such as intervention frequency or audit trail metrics.

### 6.6. Alternative Explanations

Several alternative explanations must be considered:

- **Learning Curve Effects:** Improvements may reflect learning curve dynamics rather than the effects of micro-agents alone. However, sustained improvements in assessment scores suggest that reductions in time-on-task are not solely attributable to navigation familiarity.
- **Organizational Expansion:** Expansion from 35 to 50 outlets introduced workforce heterogeneity. Such expansion would typically increase variability; the absence of such volatility suggests resilience but does not allow causal inference.
- **Novelty Effects:** The pilot phase may have been influenced by novelty effects. However, the persistence of improvements beyond the pilot phase suggests that the observed effects are not solely attributable to novelty.

These considerations reinforce the interpretation of findings as expected within the constraints of the research design.

### 6.7. Theoretical Contributions

This research contributes to several theoretical domains:

- It extends technology acceptance theory by demonstrating the influence of deployment maturity on normalization and efficiency outcomes (Davis, 1989; Venkatesh & Davis, 2000).
- It contributes to training research by employing objective behavioral measures rather than perceptual self-reports (Baldwin & Ford, 1988; Podsakoff *et al.*, 2003).
- By conceptualizing AI systems as collections of micro-agents within governance structures, it advances human-AI research beyond the dichotomy of automation and autonomy, suggesting that constrained micro-agents support sustainable performance patterns.

### 6.8. Practical Implications

From a practical standpoint, normalization should be expected and accepted rather than resisted. Perfect completion rates observed during pilot phases are unlikely to be sustained during large-scale deployment.

More importantly, the moderate efficiency effect ( $d = -0.52$ ) indicates that AI-driven micro-agents can substantially reduce time costs per unit of learning, thereby enhancing scalability. However, efficiency gains must be balanced with sustained governance to preserve learning integrity. As systems accelerate processes, the need for oversight increases rather than diminishes.

## 7. Limitations and Future Research

This research provides a longitudinal, behavioral-based approach to AI-supported training of micro-agents within an organizational context. Nevertheless, there are limitations to this research, and these limitations provide a framework for future research.

### 7.1. Contextual and Generalizability Limitations

One of the primary limitations is that, despite its focus on a rapidly expanding Italian casual dining restaurant chain, the research is specific to a particular context. Although Toscano is representative of a common frontline-intensive service environment, there is a possibility that the results of this research may not be generalizable across other industry contexts, especially those with dissimilar workforce structures, regulatory requirements, and learning environments.

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. Although the longitudinal approach reflects real-world dynamics, expansion may still affect learning outcomes. As such, future research should focus on multiple organizations and include contextual growth variables as moderators.

### 7.2. *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 remains limited. The observed improvements in learning efficiency and assessment outcomes are therefore described as statistically significant associations rather than causal effects of AI-based micro-agents.

Potential confounding variables include:

- Changes in workforce composition
- Variability in tenure
- Variability in management enforcement
- Growth of organizational units

Future studies could employ quasi-experimental methods, staggered rollouts, or feature toggling of micro-agents to isolate specific behavioral mechanisms.

### 7.3. *Hawthorne and Novelty Effects*

Thirdly, the 2024 pilot phase may have been subject to Hawthorne or novelty effects, where employees participating in the doctoral study may have demonstrated higher engagement due to increased visibility, perceived evaluation, or the novelty of the digital tool.

Although sustained improvements in efficiency observed during the 2025 phase reduce the likelihood of novelty effects, it is not possible to rule them out entirely in the 2024 phase.

Future studies could incorporate baseline or delayed-treatment groups to help disentangle novelty effects from behavioral adaptation.

### 7.4. *Platform Dependency and System-Specific Effects*

Fourth, this study was based entirely on the SafetyCulture AI-based training platform. While this platform represents a constrained, workflow-based AI micro-agent system, platform-specific effects may have contributed to the observed outcomes. Future research should compare multiple AI-based training systems to better understand how differences in micro-agent design, transparency, configuration, and autonomy levels influence adoption trajectories and efficiency gains.

### 7.5. *Measurement and Operational Definitions*

Several methodological issues require clarification:

- **Active Employee Definition:** Active employees were defined as individuals who demonstrated continued learning engagement within each phase of the study, specifically those with at least one course completion and platform interaction. This approach was intended to reduce attrition and survivor bias. However, alternative definitions based on thresholds of minimum exposure or tenure may yield different results. Future research should conduct sensitivity analyses across different thresholds.
- **Record-Level vs. Employee-Level Analysis:** While inferential analysis was conducted at the record level, descriptive statistics were reported at the employee level. Completion rate and time-on-task calculations represent aggregated course-level data. Future research should consider multilevel modeling approaches, nesting course records within employees and employees within outlets.

### 7.6. *Distributional Assumptions and Statistical Scope*

Although inferential tests were performed, distributional diagnostics such as skewness, kurtosis, and normality tests were not incorporated into the regression framework. Future research should include:

- Normality tests (e.g., Shapiro–Wilk)
- Log transformations for time variables
- Heteroskedasticity tests
- Robustness checks using non-parametric methods

### 7.7. *Temporal Scope and System Maturation*

Although the current research covers pilot and scaled phases over two years, it does not capture long-term post-adoption system maturation. Information systems research emphasizes the importance of post-adoption and routinization behaviors, which evolve over extended periods and may include habituation, oversight recalibration, efficiency plateauing, and engagement decay.

Long-term longitudinal research would be valuable for examining:

- Habituation
- Oversight recalibration
- Efficiency plateauing
- Engagement decay

These represent important directions for future research on post-adoption information systems maturation.

### 7.8. Future Research Agenda

Based on the limitations identified, future research should focus on the following directions:

1. **Experimental Micro-Agent Manipulation** Random activation and deactivation of specific micro-agent features (e.g., nudging, sequencing, reinforcement).
2. **Governance Intensity Modeling** Quantification of oversight and inclusion in regression models to test moderation effects.
3. **Contextual Growth Moderation** Inclusion of organizational expansion variables in regression models to examine moderation effects.
4. **Cross-Platform Comparative Research** Examination of how differences in AI-enabled training architectures influence longitudinal adoption and efficiency patterns.

## 8. Conclusion

As artificial intelligence becomes more integrated into organizational learning systems, it is expected that evaluation processes will move beyond pilot-based snapshots toward longitudinal data based on actual operating conditions. This research paper contributes to this progression by offering behavior-based longitudinal evidence for AI-enabled training micro-agents in an organization transitioning from pilot to scale within a frontline-intensive context.

The results indicate that, in the transition from pilot to scale, there was:

- Improved completion rates
- Improved learning efficiency with a moderate effect size
- Improved assessment results
- Stable adoption even in the context of organizational expansion

These results suggest that AI micro-agents, operating in a constrained state within human-governed systems, are associated with sustained behavioral adaptation over time. However, it should be noted that the research paper does not suggest that the efficiency gains were directly attributable to the AI micro-agents themselves. Rather, it suggests that, within a human-in-the-loop governance structure, AI-enabled training was associated with improved fluency and mastery at scale.

From a conceptual perspective, this research offers a significant contribution by advancing a micro-agent-based understanding of AI-supported training that is task-bound, workflow-integrated, and behavioral in nature, shaping learning processes without replacing human agency. This perspective provides a middle ground between technological optimism in automation and skepticism regarding governance.

From a theoretical perspective, this research contributes to advancing:

- Post-adoption IS maturation theory

- Training transfer and reinforcement theory
- Human–AI complementarity theory

From a methodological perspective, this research contributes through the use of platform-derived learning trace data, which provides objective insights into longitudinal effects while reducing reliance on self-reported measures.

Practically, this research highlights that AI-supported training systems should be designed with normalization in mind rather than perfection in pilot outcomes. Governance structures play a critical role in achieving efficiency gains.

In conclusion, AI-supported training systems hold significant promise not in terms of their intelligence per se, but in terms of the responsible constraint of micro-agents within accountable human systems. By moving beyond pilot-based evaluation, this research offers a replicable model for organizations seeking to integrate AI into their training systems.

## Acknowledgement

The author declares that there are no acknowledgments for this research paper.

## Authorship Contribution

The authors solely contributed to the conceptualization and design of the study, data collection, data analysis, interpretation of results, and preparation of the manuscript.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declarations

The authors confirm that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere. All sources of information have been properly acknowledged, and ethical research standards have been followed.

## Ethical Approval

The study involved analysis of anonymized, system-generated learning data collected as part of routine organizational training. No personal identifiers were used, and no human subjects intervention was conducted.

## Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Data Availability Statement

Data were generated from an AI-enabled training platform used in an organizational setting. Aggregated, anonymized data may be made available upon reasonable request, subject to organizational and ethical constraints.

## AI Usage Statement

The author declares the use of generative AI tools, including writing assistants and grammar refinement systems, to support manuscript organization, drafting, and language improvement. These tools were not used for data generation, statistical analysis, or interpretation of results.

## References

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* 1-13. <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. <https://doi.org/10.1037/xge0000033>
- DotePardo, J.S.(2025). Artificial intelligence in organizational learning: a systematic review of applications, opportunities, and challenges. *Development and Learning in Organizations: An International Journal*, 1-6. <https://doi.org/10.1108/DLO-06-2025-0228>
- Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2019). Cognitive challenges in human-AI collaboration: Investigating the path towards productive delegation (SSRN Scholarly Paper 3368813). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3368813>
- Hemmer, P., Schemmer, M., Köhl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human-AI collaboration: Concept, sources, and evidence. *European Journal of Information Systems*, 34(6), 979-1002. <https://doi.org/10.1080/0960085X.2025.2475962>
- Muehleman, 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. *The Journal of Strategic Information Systems*, 34(2), 101885. <https://doi.org/10.1016/j.jsis.2024.101885>
- Pappas, I. O., Giannakos, M. N., & Sampson, D. G. (2019). Fuzzy set analysis as a means to understand users of 21st-century learning systems: The case of mobile learning and reflections on learning analytics research. *Computers in Human Behavior*, 92, 646-659. <https://doi.org/10.1016/j.chb.2017.10.010>
- 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: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879. <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: Four longitudinal field studies. *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<sup>1</sup>. *MIS quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology<sup>1</sup>. *MIS quarterly*, 36(1), 157-178. <https://doi.org/10.2307/41410412>