

In this issue:

- 4. CurriculumGPT: Supporting Curriculum Design with AI-Augmented Tooling**
Kareem Dana, West Texas A&M University
Jeffry Babb, West Texas A&M University

- 23. A Comparative Case Study of AI-Enabled and Traditional Experiential Simulations in Operations and Supply Chain Education**
Yan Jin, Quinnipiac University
Ae-Sook Kim, Quinnipiac University
Tan Gürpınar, Quinnipiac University
Idrisu Awudu, Quinnipiac University

- 37. Personality, Organizational Environment, and the Use of Generative AI: A Conceptual Framework for Understanding Human-AI Interaction**
David Firth, University of Montana
Simon Laub, Business Academy Aarhus
Elizabeth Kohl, University of Montana
Fang Chen, University of Montana

- 55. Adapting CAMIL for K–12 World Language Learning: A Qualitative Study of Immersive VR in Middle School Spanish**
Emily Loughlin, Nichols College
Kevin Mentzer, Nichols College

- 66. Developing an Artificial Intelligence Course for a Small Undergraduate Program**
Scott Turner, University of Richmond
Lionel Mew, University of Richmond
Terry Turner, University of Richmond

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is five times per year. The first year of publication was 2003.

ISEDJ is published online (<https://isedj.org>). Our sister publication, the Proceedings of the ISCAP Conference (<https://iscap.us/proceedings>) features all papers, abstracts, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the ISCAP conference. All papers, whether award-winners or not, are invited to resubmit for journal consideration after applying feedback from the Conference presentation. Award winning papers are assured of a publication slot; however, all re-submitted papers including award winners are subjected to a second round of three blind peer reviews to improve quality and make final accept/reject decisions. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 35%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to volunteer members of ISCAP who perform the editorial and review processes for ISEDJ.

2026 ISCAP Board of Directors

Amy Connolly
James Madison University
President

Michael Smith
Georgia Institute of Technology
Vice President

Jeff Cummings
Univ of NC Wilmington
Past President

David Firth
University of Montana
Director

Mark Frydenberg
Bentley University
Director/Secretary

Leigh Mutchler
James Madison University
Director

RJ Podeschi
Millikin University
Director/Treasurer

Bryan Reinicke
Rochester Institute of
Technology / Director

Jeffry Babb
West Texas A&M University
Director/Curricular Matters

Eric Breimer
Siena University
Director/2026 Conf Chair

Thomas Janicki
Univ of NC Wilmington
Director/Meeting Planner

Xihui "Paul" Zhang
University of North Alabama
Director/JISE Editor

Copyright © 2026 by Information Systems and Computing Academic Professionals (ISCAP). Permission to make digital or hard copies of all or part of this journal for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial use. All copies must bear this notice and full citation. Permission from the Editor is required to post to servers, redistribute to lists, or utilize in a for-profit or commercial use. Permission requests should be sent to Kevin Mentzer, Editor, editor@isedj.org.

INFORMATION SYSTEMS EDUCATION JOURNAL

Editors

Kevin Mentzer
Editor
Nichols College

Ira Goldman
Associate Editor
Siena University

David Yates
Associate Editor
Bentley University

Michelle Louch
Teaching Cases & Exercises
Editor
University of Pittsburgh - Greensburg

Mark Pisano
Teaching Cases & Exercises
Associate Editor
Southern Connecticut
State University

Thomas Janicki
Publisher
Univ of NC Wilmington

David Woods
Assistant Publisher
Miami University
Regionals

Paul Witman
Emeritus Editor
(2021-2026)
California Lutheran
University

Jeffry Babb
Emeritus Editor
(2016-2021)
West Texas A&M
University

Donald Colton
Emeritus Editor
(2003-2010)
Brigham Young University
Hawaii

CurriculumGPT: Supporting Curriculum Design with AI-Augmented Tooling

Kareem Dana
kdana@wtamu.edu

Jeffrey Babb
jbabb@wtamu.edu

West Texas A&M University
Canyon, TX 79016, USA

Abstract

A key challenge in model curriculum development, especially for information systems, is its lengthy revision cycle, often spanning 5 to 10 years, and static reports that quickly become outdated. Recognizing this, recent model curricula, CC2020 and IS2020, call for a more sustained and continuous development process. In response, we present CurriculumGPT, a prototype AI tool built with a retrieval-augmented generation (RAG) architecture, to help realize that goal and address limitations of a prior tooling effort, the Computing Competencies Curricula Tool. We conducted a pilot study to evaluate the system across representative curriculum design tasks, including fact retrieval, comparative analysis, and content creation. Our training corpus included syllabi, course materials, learning objectives, and model curricula reports. Responses were assessed with a standardized human-applied rubric and the automated RAG evaluation framework, RAGAS, which evaluated quantitative metrics of answer faithfulness, relevancy, context precision, and recall. Results indicate that CurriculumGPT performed reliably, with improved performance achieved through metadata-aware chunking, a refinement to standard RAG approaches that preserves structural relationships in the data. Our findings show how AI and RAG systems can be adapted to support curriculum design while highlighting limitations and challenges that establish a foundation for future research in this emerging area.

Keywords: Curriculum Design, Generative AI, Computing Education, Competency, Curriculum Tooling, AI in Education

Recommended Citation: Dana, K., Babb, J., (2026). CurriculumGPT: Supporting Curriculum Design with AI-Augmented Tooling. *Information Systems Education Journal*, v25(n1) pp 4-22. DOI# <https://doi.org/10.62273/CVEC6056>

CurriculumGPT: Supporting Curriculum Design with AI-Augmented Tooling

Kareem Dana and Jeffrey Babb

1. INTRODUCTION

Computing education continually evolves, shaped by emerging technologies and industry demands. As the field advances, so does the debate over what graduates should know and be able to do. As a result, developing model curricula in computing and information systems is both essential and inherently challenging.

Recent initiatives to guide this process include the Computing Curricula 2020 (CC2020) and IS2020 reports (Clear et al., 2020; Leidig et al., 2020). These reports strongly advocate for a move towards a competency-based computing curriculum which emphasizes what graduates should be able to do rather than just what they know (Babb et al., 2025). Both reports, however, are ultimately static documents produced through lengthy development cycles.

Recognizing this limitation, both CC2020 and IS2020 call for a more continuous, community-driven approach supported by effective software tooling. IS2020 envisions a living document community where educators collaboratively update curricula over time (Babb et al., 2021; Clear et al., 2020; Leidig et al., 2020).

Initial efforts to realize this vision included the Computing Competencies Curricula Tool or C3T, developed through a collaboration between the Information Systems & Computing Academic Professionals (ISCAP) and the Association for Computing Machinery (ACM). C3T demonstrated the feasibility of a web-based platform for creating and comparing competencies. However, it relied heavily on manual data entry, offered limited interactivity, and contained only a small initial dataset. These limitations constrained its ability to support scalable and continuous curriculum development (C3T, 2025).

This highlights a broader problem. There is currently no effective tooling that supports dynamic, data-driven curriculum design by leveraging the large volume of unstructured curricular artifacts already available to educators, such as syllabi, course catalogs, and learning objectives. Recent advances in generative AI, particularly large language models (LLMs) and

retrieval-augmented generation (RAG) offer a promising foundation to address this challenge.

Following a design science approach, we developed and evaluated CurriculumGPT, an AI-augmented curriculum design tool built on a RAG architecture tailored to the structure of curricular data. In contrast to generic RAG implementations, our approach incorporates metadata-aware chunking to preserve relationships within curricular data, improving retrieval fidelity. The system supports natural language interaction enabling educators to query, analyze, and generate curriculum content more efficiently, while also providing source attribution to support transparency and explainability of responses.

This paper makes three contributions:

- First, we present CurriculumGPT, our AI tool, which applies a RAG architecture to support curriculum development tasks using academic data.
- Second, we introduce and evaluate metadata-aware chunking as a refinement for RAG systems operating on hierarchical curricular data, demonstrating its impact on retrieval accuracy and task performance.
- Third, we provide initial design insights into the use of AI-powered tools for curriculum design, including considerations for transparency, explainability, and human oversight, highlighting both their potential and limitations through an exploratory pilot study, and discuss future research paths in this area.

Together, these contributions extend prior work on curriculum tooling by demonstrating how AI can be adapted to the specific challenges of curriculum design, while also contributing to emerging design knowledge on the application of RAG in structured, domain-specific contexts.

2. LITERATURE REVIEW

IS2020 (Leidig et al., 2020) defines a competency as “the graduate’s ability to apply knowledge, skills, and dispositions to effectively complete

tasks.” In this definition, knowledge is the most familiar component, referring to factual elements embedded in syllabi and course materials that students should know. Skills represent the ability to apply that knowledge in practice, often aligned with Bloom’s taxonomy to capture cognitive complexity (e.g., remember, analyze, create). Dispositions encompass the soft skills such as ethics, professionalism, and values that guide how knowledge and skills are applied in context (Babb et al., 2025; Leidig et al., 2020). To support software tooling, Babb et al. (2020) introduced the eXtensible Computing Curriculum Reporting Language (XCCRL), a structured data format for representing competencies and their components. Building on this foundation, C3T was developed as a web-based platform to help educators build competencies (C3T, 2025).

Much of the recent literature on AI in education has focused on enhancing the student learning experience. Marquardson (2024) explored how AI tools can support self-directed learning, while Mew (2024) examined the pedagogical role of ChatGPT in higher education. Lang and Gürpınar (2025) built a RAG chatbot to enhance student learning in an online course. Dickey & Bejarano (2024) developed GAIDE, a generative AI-powered assistant designed to help instructors draft course content and generate teaching materials.

Other studies have begun exploring how generative AI can support curriculum development itself. Padovano & Cardamone (2024) proposed a structured approach for using AI to define engineering competencies, while Sridhar et al. (2023) conducted an empirical study on the effectiveness of large language models (LLMs) in generating course-level learning objectives. Iweuno et al. (2024) discuss using AI to improve curriculum design by enabling data-driven decision making, personalized learning, and more inclusive curricular content. Lastly, Chai et al. (2025) performed a thorough review of other generative AI uses in instructional design. These studies suggest that AI can play a meaningful role in instruction and curriculum design. However, to our knowledge no previous work has explicitly developed and empirically evaluated AI tools with a focus on IS curricula.

Finally, Eduglopedia, though now defunct, represented an early attempt to create a global, open encyclopedia of educational programs and curricular resources through a collaborative web-based platform. It built an impressive collection of content that led to the release of annual reports, which shared trends, insights, and

guidance with the IS community. This effort demonstrated the need and interest in software support for curriculum design (vom Brocke et al., 2020).

3. CASE STUDY: THE C3T TOOL

Purpose and Vision of C3T

One of the key ideas emphasized in the CC2020 and IS2020 reports is that curriculum development should be a sustained and ongoing process, not a one-time event every few years (Clear et al., 2020; Leidig et al., 2020). These reports, along with the work of Babb et al. (2021), call for moving away from traditional, static curriculum models in favor of more flexible and dynamic models. C3T was developed as an initial step towards realizing this vision. Its primary purpose was to offer educators and other stakeholders a web-based platform through which they can build, visualize, and compare competencies (C3T, 2025).

A secondary, yet more ambitious goal was to serve as a community-supported platform for the collective stewardship of model curricula (C3T, 2025). With such a community, C3T could realize the vision of a continuous model curriculum that adapts quickly to emerging innovations. For example, the topic of artificial intelligence is not highlighted explicitly as a competency in the IS2020 report. With a living document community in place, educators could have already collaboratively proposed, reviewed, and integrated AI-related competencies. These contributions could then be easily shared across universities well before publication of the next model curriculum.

Architectural Overview

C3T was developed as a software-as-a-service application to support the creation, management, and comparison of curriculum competencies. Development occurred incrementally between 2019 and 2024 through the part-time work of IS faculty and student developers (C3T, 2025).

Early development focused on basic competency management, including create, read, update, and delete functionality. Subsequent enhancements added search, filtering, and table-based comparison features (C3T, 2025).

A later development phase envisioned C3T as a community-driven platform incorporating collaboration, role-based governance, and discussion forums. Preliminary work was completed on user registration. However, many of the critical community features remain

incomplete (C3T, 2025).

Outcome and Lessons Learned

In November 2024, C3T was publicly deployed and presented at an academic conference. Development and deployment yielded several important insights. First, the tool succeeded in demonstrating the feasibility of a web-based platform for modeling computing competencies. However, since its deployment, user engagement has been minimal. While the website received visitors, there is no evidence of significant user registration, collaboration, or competency creation (C3T, 2025).

Several factors appear to contribute to this lack of adoption. First, the process of entering a single competency involves multiple steps and decisions, with significant manual data entry. This is a high effort process that likely discourages visitors. Second, C3T contains little existing data. Like a new social media platform, it lacks the critical mass of content needed to make participation worthwhile. Third, the user interface is minimal and lacks competency visualizations or interactive elements to help users interpret, compare, and make decisions about curricula. Finally, the living document features have not been implemented, limiting C3T's potential to support such a community. These limitations underscore our motivation to explore AI-augmented curriculum tooling as the next stage of tooling development.

4. CURRICULUMGPT

Drawing on the limitations of C3T and the characteristics of curricular data, we developed CurriculumGPT with four key design goals:

1. CurriculumGPT should intelligently analyze documents such as syllabi, course catalogs, and learning objectives to automatically extract relevant content such as competencies, reducing the need for manual data entry.
2. Users should be able to interact with CurriculumGPT using conversational questions such as "Which competencies align with this course?" or "What's missing from our cybersecurity curriculum?". Such interactions can increase engagement and enable more informed decision making.
3. The tool should preserve the structure within curricular data. Such data contains important relationships across courses, learning objectives, and competencies, which must be maintained during retrieval.

4. Users should be able to trace responses back to source materials to assess accuracy and build trust.

Architecture Overview

We developed a prototype implementation of CurriculumGPT based on these design goals. Our architecture is grounded in design science research (Hevner et al., 2004), which emphasizes building artifacts to address identified problems and evaluating their effectiveness, and information retrieval principles, particularly the RAG paradigm (Lewis et al., 2020).

Our prototype is built using Streamlit, LlamaIndex, and OpenAI's GPT-4o LLM. Streamlit is a Python framework that provides easy-to-use web interfaces for users and easily integrates with LLMs. See Appendix A for a screenshot of CurriculumGPT that illustrates Streamlit and the user interface.

LlamaIndex is the RAG pipeline used in our prototype. RAG allows the system to generate responses based on curated academic documents, rather than relying solely on the LLM's general pretraining (Lewis et al., 2020). LlamaIndex searches the curated documents and passes the most relevant chunks to the LLM. This addresses a critical challenge in using general-purpose LLMs for education tooling: lack of contextual awareness about specific curricula or local academic data (Lang & Gürpinar, 2025). Our prototype includes explainability features by citing the specific documents retrieved from the training data during each query. Finally, the LLM used is GPT-4o. See Figure 1 below for a diagram of the CurriculumGPT architecture.

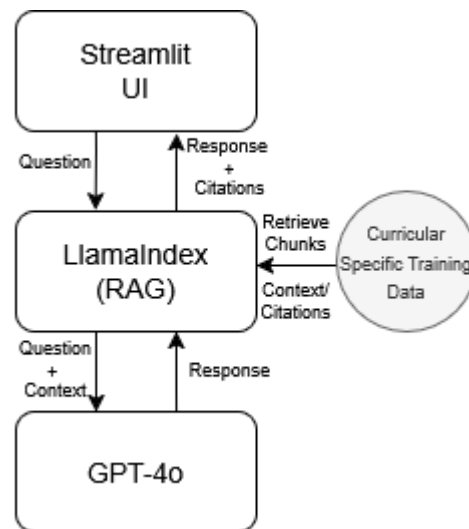


Figure 1: CurriculumGPT Architecture

Type of Data	Number of Documents (43 documents)	Used in CurriculumGPT	Purpose for Inclusion
Model Curricula	2 Documents (136,461 words)	IS2020, CC2020	Grounds responses in standards; ensure alignment with models.
Course Materials	36 documents (129,103 words)	Course catalogs, degree plans, syllabi, Course Learning Objectives (CLOs)	Supports queries about specific courses, course comparisons, and demonstrates how courses are structured.
Accreditation standards	1 document (6,614 words)	ABET criteria	Supports accreditation tasks; helps faculty align content with accreditation requirements.
Competency Frameworks	3 documents (4,988 words)	Matrices mapping skills, knowledge elements, dispositions to competencies	Structures competency relationships; enables comparisons and gap analysis.
Taxonomies	1 document (384 words)	Bloom’s Taxonomy	Guides phrasing of objectives at appropriate cognitive levels.

Table 1: Training Data Used in CurriculumGPT

While RAG has been widely applied in many domains, our architecture adapted and refined it for curriculum design. Curricular data is inherently hierarchical and relational, with dependencies across courses, learning objectives, and competencies that are not well suited to naive chunking or retrieval. To preserve those relationships and improve retrieval accuracy, we implemented metadata-aware chunking, tagging each chunk with its associated course name or relationship (Danter et al., 2024; Qu et al., 2024). The RAG architecture also supported the implementation of source tracing and updates to the training corpus without requiring model retraining.

Implementation Details

To support reproducibility and provide greater transparency of our system, key technical aspects are described below.

The training corpus consisted of curricular materials from a single academic department, including course syllabi, course catalogs, degree plans, and course learning objectives. It also included model curriculum reports (IS2020 and CC2020), as well as competency frameworks and taxonomies (e.g., Bloom’s Taxonomy). In total, the corpus included forty-three documents. Refer to Table 1 for the complete training data details.

Documents were segmented using a sentence-based chunking strategy. Each document was divided into chunks of approximately 1024 tokens, with an overlap of 200 tokens between adjacent chunks. A token is a word or part of a

word used by language models during processing and generation. To improve retrieval accuracy, a metadata-aware chunking mechanism was implemented. Course and hierarchy identifiers were embedded directly into each chunk, ensuring semantic continuity across chunks.

Chunk retrieval was performed using a vector similarity search with the top 10 chunks selected per query. Additionally, a curriculum domain-specific re-ranking step was applied as part of the metadata-aware chunking technique to prioritize chunks whose metadata matched the requested curricular component (e.g., course or learning objective). This hybrid retrieval approach helped improve precision.

Responses were generated using the OpenAI GPT-4o model. A default temperature setting of 0.5 was used. The model was prompted with both retrieved documents and a master prompt guiding it to prioritize relevant, source-based responses while leveraging its general LLM knowledge when necessary.

The source code for CurriculumGPT is available at: <https://github.com/kareemy/CurriculumGPT/>

5. EVALUATION AND RESULTS

To evaluate CurriculumGPT, we designed a series of queries across three categories of tasks: basic retrieval, comparative analysis, and content creation. Queries were developed to represent realistic tasks faced by curriculum designers, based on examples drawn from faculty

Task Category	Faithfulness w/o and with		Response Relevancy		Context Recall		Context Precision	
	Retrieval	0.79	0.96	0.92	0.96	0.73	0.97	0.64
Comparison	0.46	0.82	0.87	0.89	0.61	0.83	0.45	0.71
Creation	0.76	0.79	0.94	0.94	N/A		0.61	0.78

Table 2: Average RAGAS Evaluation Metrics without and with Metadata-aware Chunking

Task Category	Faithfulness (Accuracy)	Relevance	Clarity	Actionability
Retrieval	4.50	4.67	4.58	4.67
Comparison	3.50	3.83	4.00	4.00
Creation	3.83	4.50	3.75	3.50

Table 3: Average Reviewer Rubric Scores (1 to 5)

experiences and model curriculum guidelines. Twenty queries and responses from each category (60 total) were evaluated using RAGAS, a standardized, automated framework for evaluating RAG+LLM systems across four quantitative metrics: faithfulness, response relevancy, context recall, and context precision (each scored 0-1).

Faithfulness measures how factually consistent a response is with the training data. A score of 1 indicates that the response is fully supported by the training data, while lower scores suggest hallucinations or that the tool is answering based on its general LLM knowledge instead of the curriculum-specific corpus. Response relevancy measures how directly and appropriately the response addresses the query. RAGAS measures this by reverse-engineering the question from the response and comparing it to the original, with lower scores indicating incomplete or off-topic answers. (Es et al., 2024)

Context recall measures how well CurriculumGPT retrieves and incorporates relevant information from the training data. RAGAS performs this evaluation by comparing the response to a correct, reference response. A low score suggests that the response is missing important information. Finally, context precision measures the proportion of retrieved chunks that were relevant. Low precision means many irrelevant or unused chunks were retrieved. These metrics are established measures, with full methodological details documented on the RAGAS website and supporting publications (Es et al., 2024).

The results with and without metadata-aware chunking are summarized in Table 2. Lower scores (left) were without metadata-aware chunking. In addition to automated metrics, four independent reviewers evaluated the responses using a standardized rubric (Appendix B). The

rubric scored responses on a 1-5 Likert scale for faithfulness, relevance, clarity, and actionability. Appendix C includes queries and responses that are representative of each category.

Reviewers were faculty in our IS department and selected based on their familiarity with curriculum design. Reviewers were given guidance on how to interpret each dimension. To assess the consistency of reviewer evaluations, we calculated inter-rater reliability using the intraclass correlation coefficient, ICC(2,k), for each evaluation dimension. Results indicated good agreement for faithfulness (ICC = 0.88), relevance (ICC = 0.84), and actionability (ICC = 0.82), and moderate agreement for clarity (ICC = 0.74), suggesting that the rubric was applied consistently across reviewers.

System Validation: Retrieval and Fact Recall

In this first category, CurriculumGPT was prompted to retrieve specific facts from the training data. It performed reliably. It was able to retrieve accurate course learning objectives and descriptions from syllabi. It was also able to correctly define important curricular terms such as “knowledge-skill pair” from the appropriate source material. As shown in Table 2, this category achieved the highest overall scores across the RAGAS metrics, averaging above 0.9 with metadata-aware chunking. Reviewers also rated these responses highly on the rubric.

Comparisons: Syllabus Gap Analysis and Competency Comparison

This category involved more complex tasks, including asking CurriculumGPT to identify competency coverage gaps across courses and compare content. CurriculumGPT scored lower across both the RAGAS metrics and reviewer evaluations compared to the other categories. These tasks were most prone to hallucinations, but also most improved by metadata-aware

chunking. Future technical tweaks and a larger training corpus could improve these types of responses even further.

Creation: Authoring Competencies and Course Outlines

In tasks asking CurriculumGPT to create new content, it performed relatively well. The faithfulness metric was naturally lower since the tool generated content not explicitly present in the training data and context recall was not measured because these tasks did not have a definitive reference answer. In this case, the response relevancy metric and reviewer scores were the best indicators of usefulness. The system produced plausible draft competencies and outlines for new courses (high response relevancy), demonstrating its utility as a brainstorming aid or starting point for curriculum planning. However, reviewers indicated its output often created broad competencies containing multiple learning objectives that were more complex than necessary. Improved prompting and follow-up queries can help refine these results. For example, follow-up queries such as *"Separate this competency into three distinct competencies."* or *"Revise using Bloom's taxonomy at the analyze level."* led to improved results.

Benefits & Faculty Use Cases

One of the most immediate use cases of CurriculumGPT is its ability to create course-level competencies and learning objectives. Faculty can use its responses as a starting point, but refinement will probably be needed to produce clear, appropriately scoped competencies. These results are consistent with the research of Sridhar et al. (2023) who also studied whether LLMs can generate learning objectives.

We also found that our prototype may be beneficial in helping faculty explore ideas for new courses or revise existing ones. Leveraging its training data, faculty can ask questions such as, *"Generate a sample syllabus outline for a course that teaches cybersecurity and maps to the CC2020 competencies."* or *"I'm revising our course on Data Analytics. Based on current industry trends and existing competencies in our program, what new topics or learning objectives should I consider adding?"* This result is consistent with recent research on the topic of using LLMs for curriculum design (Iweuno et al., 2024; Padovano & Cardamone, 2024).

Overall, these results demonstrate that CurriculumGPT shows promise as a curriculum support tool. However, it has limitations,

challenges remain, and future work is needed.

6. LIMITATIONS AND FUTURE WORK

Limitations

This study represents an initial, exploratory evaluation of CurriculumGPT and contains several limitations. First, the system was evaluated using curricular data from a single department within one university. This narrow scope limits the generalizability of the findings and does not fully capture how the system would perform across different universities, disciplines, or curriculum structures. While the results provide a proof of concept and highlight promising architectural design choices, future work should examine system performance using larger and more diverse datasets.

Second, the evaluation relies, in part, on RAGAS metrics. These metrics evaluate text generation quality, not curriculum design or educational quality. As a result, high RAGAS scores should not be interpreted as evidence of improved curriculum design outcomes. Instead, they indicate that responses are well-grounded in retrieved source material and aligned with the input query. This distinction is important, as effective curriculum design requires domain expertise and holistic judgment beyond what automated metrics can capture.

Third, four human reviewers were included to complement the RAGAS evaluation. While inter-rater reliability analysis indicated moderate to good agreement, a larger and more diverse group of reviewers would provide a more robust assessment of the system.

Finally, while we compared standard RAG with metadata-aware chunking, we did not include a baseline using an unguided LLM without RAG, nor comparisons to existing curriculum design tools or human-only workflows. As a result, the findings should be interpreted as evidence of relative improvement within the RAG-based approach rather than a comprehensive evaluation of alternative methods. Future research could incorporate more comparisons to better assess the effectiveness of AI-augmented curriculum design tools.

Future Work

The results of this pilot study reveal several opportunities to improve CurriculumGPT, particularly in supporting more complex curriculum tasks such as comparison and gap analysis.

Knowledge graphs provide a way to keep the training data interconnected and maintain relationships between courses, learning objectives, and competencies. Incorporating knowledge graphs to explicitly model curricular relationships could enable the system to better respond to those more complex queries (Ji et al., 2021).

Additionally, integrating table-augmented generation (TAG) or text-to-SQL capabilities could further improve the tool by allowing it to interface directly with structured databases (Biswal et al., 2024). Together, these extensions point to the need for a mechanism to coordinate how structured and unstructured curricular data can be accessed within the system.

Another promising direction is the Model Context Protocol (MCP), an open standard that allows LLMs to interact with external data systems (Karimova & Dadashova, 2025). Adopting MCP could allow CurriculumGPT to interface with both knowledge graphs and SQL databases through a single, standardized protocol. This architecture could also lay the groundwork for agentic workflows, where the tool can autonomously query live data or execute complex analysis tasks. Similar MCP-based approaches have already been demonstrated in the structural engineering domain (Avila et al., 2025).

Improving explainability remains an important area for future work. We implemented source tracing and inline citations within CurriculumGPT, but more explainability is needed. Exploring toolkits like TruLens (TruLens, n.d.) could be valuable. TruLens integrates within the RAG pipeline to validate retrieval accuracy and offer interactive explanations to help users trace how the system arrived at a response.

We also see several other applications for CurriculumGPT or AI curriculum tools that can be explored in future work. One is its use as a support tool in accreditation-related tasks, such as ABET or AACSB reviews. When trained on accreditation criteria and an institution's courses, CurriculumGPT could help map course content to accreditation criteria or synthesize data needed for a report. Another potential application is training the tool on job postings to extract desired skills and trends in industry, which could inform and align model curricula with workforce expectations.

Adding these capabilities would be an important step toward realizing the living document community vision as outlined in IS2020. While

CurriculumGPT is a first step, that vision remains aspirational and will depend on this future work.

7. AI RISKS AND GOVERNANCE

This section focuses on the broader implications of a tool like CurriculumGPT, highlighting risks and governance practices we believe are necessary to ensure responsible and successful use of AI in curriculum design.

Using an LLM in educational settings introduces several ethical and institutional risks that must be considered carefully. At the heart of that discussion is the training data. An LLM's ability to generate accurate and domain-specific responses depends on the quality of the training data (Lang & Gürpınar, 2025; Lin et al., 2024).

Decisions about what training data to include can greatly impact the quality of the system's output and introduce systemic bias or blind spots if not carefully reviewed (Ladwig & Schwieger, 2024; Lin et al., 2024). For example, an LLM could suggest outdated terminology, reinforce gender stereotypes, or fail to account for regional legal and cultural differences in topics like cybersecurity or ethics.

There are also privacy concerns associated with indexing internal university documents, particularly if those documents contain sensitive information. Likewise, LLMs may be pre-trained on copyrighted material, raising legal and ethical questions about reusing that content in applications like CurriculumGPT (Ladwig & Schwieger, 2024).

Another major concern is AI hallucination, the generation of plausible but inaccurate information, which could mislead users relying on the tool for curriculum decisions (Ladwig & Schwieger, 2024; Lin et al., 2024).

Future development of AI curriculum tools should proceed with a clear governance structure that focuses on ethical design and human oversight. AI tools to improve productivity are here to stay and they will find their way into curriculum design, either by us or others. Only with human oversight can these risks be properly managed.

Prior work, including IS2020 (Leidig et al., 2020) and Babb et al. (2021), outlines potential governance structures that include a board of directors and various task forces. We support these models and suggest they add a dedicated AI oversight representative or task force to ensure that AI is used responsibly and trained on

a diverse set of data that accurately reflects the will of the IS community.

8. CONCLUSIONS

This paper presents CurriculumGPT as an initial step towards supporting curriculum design with AI-augmented tooling. Our evaluation showed that the tool performed reliably on retrieval tasks, achieving high scores across all four RAGAS metrics (above 0.9) and strong reviewer ratings (above 4.5/5.0). Comparative tasks, such as identifying competency gaps across courses, were more challenging, but accuracy improved notably with metadata-aware chunking. For content creation, the system generated plausible draft competencies and course outlines that reviewers found useful for brainstorming, though results were noted as being too broad at times.

The introduction of metadata-aware chunking and source attribution are two key technical contributions of this work. They addressed a limitation of standard RAG approaches and illustrated how RAG architectures can be adapted to better support domain-specific tasks such as curriculum design.

Perhaps more importantly, this work establishes a foundation for future research. Exploring new, richer ways to represent curricular data and more advanced approaches to retrieval can further enhance the system's usefulness for IS educators.

Building tools like CurriculumGPT and further research contribute towards the living document vision articulated in IS2020. Realizing this vision will require continued collaboration among educators, researchers, and professional organizations to ensure that these tools are designed, used, and governed in ways that reflect the needs and values of the IS community.

9. REFERENCES

- Avila, C., Ilbay, D., & Rivera, D. (2025). Human-AI Teaming in Structural Analysis: A Model Context Protocol Approach for Explainable and Accurate Generative AI. *Buildings*, 15(17), 3190. <https://doi.org/10.3390/buildings15173190>
- Babb, J., Gardner, L., Leidig, P. M., Salmela, H., Sooriamurthi, R., Muthupoltotage, U. P., & Qutab, S. (2021). Towards a Continuous Process for Using and Sustaining IS Curriculum Guidelines. *Proceedings of the 2021 AIS SIGED International Conference on Information Systems Education and Research*. <https://aisel.aisnet.org/siged2021/3>
- Babb, J., Sharp, J. H., Waguespack, L., Abdullat, A., & Dana, K. (2020). eXtensible Computing Curriculum Reporting Language (XCCRL). *Information Systems Education Journal*, 18(6), 13-27.
- Babb, J., Yates, D., & Waguespack, L. (2025). On Becoming: Why Disposition Distinguishes Information Systems Education from Training. A Commentary on Model Curricula. *Information Systems Education Journal*, 23(1). <https://doi.org/10.62273/JRSO9158>
- Biswal, A., Patel, L., Jha, S., Kamsetty, A., Liu, S., Gonzalez, J. E., Zaharia, M. (2024). Text2sql is not enough: Unifying ai and databases with tag. arXiv preprint arXiv:2408.14717.
- C3T. (n.d.). C3T: Computing Competencies Curricula Tool. Retrieved June 22, 2025, from <https://c3t.dev/>
- Chai, D. S., Kim, H. S., Kim, K. N., Ha, Y., Shin, S. S. H., & Yoon, S. W. (2025). Generative Artificial Intelligence in Instructional System Design. *Human Resource Development Review*. <https://doi.org/10.1177/15344843251320256>
- Clear, A., Parrish, A., Ciancarini, P., Cuadros-Vargas, E., Frezza, S., Gal-Ezer, J., Impagliazzo, J., Pears, A., Takada, S., Topi, H., van der Veer, G., Vichare, A., Waguespack, L., Wang, P., Zhang, M. (2020). Computing Curricula 2020 (CC2020), *Paradigms for Future Computing Curricula*, ACM, IEEE, <https://doi.org/10.1145/3467967>
- Danter, D., Mühle, H., & Stöckl, A. (2024). Advanced Chunking and Search Methods for Improved Retrieval-Augmented Generation (RAG) System Performance in E-Learning. *Human Factors in Design, Engineering, and Computing*, 159(159). <https://doi.org/10.54941/ahfe1005756>
- Dickey, E., & Bejarano, A. (2024, October). GAIDE: A Framework for Using Generative AI to Assist in Course Content Development. In *2024 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE. <https://doi.org/10.1109/fie61694.2024.10893132>
- Es, S., James, J., Anke, L. E., & Schockaert, S. (2024, March). RAGAS: Automated evaluation of retrieval augmented generation. In *Proceedings of the 18th Conference of the European Chapter of the*

- Association for Computational Linguistics: System Demonstrations* (pp. 150-158). <https://doi.org/10.18653/v1/2024.eacl-demo.16>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2(1).
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105. <https://doi.org/10.2307/25148625>
- Iweuno, B. N., Orekha, P., Ojadiran, O., Imohimi, E., & Adu-Twum, H. T. (2024). Leveraging Artificial Intelligence for an inclusive and diversified curriculum. *World Journal of Advanced Research and Reviews*, 23(2), 1579-1590. <https://doi.org/10.30574/wjarr.2024.23.2.2440>
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2), 494-514. <https://doi.org/10.1109/tnnls.2021.3070843>
- Karimova, S., & Dadashova, U. (2025). The Model Context Protocol: A Standardization Analysis for Application Integration. *UNEC Journal of Computer Science and Digital Technologies*, 1(1), 50-59. <https://doi.org/10.30546/UNECCSDT.2025.01.050>
- Ladwig, C., Schwieger, D. (2024). An Eye Toward the Softer Side of CC2020 Computing Curricula: Professional, Legal, and Ethical Artificial Intelligence Issues. *Information Systems Education Journal*, 22(1), pp.53-88. <https://doi.org/10.62273/FFDJ1170>
- Lang, G., Gürpınar, T. (2025). AI-Powered Learning Support: A Study of Retrieval-Augmented Generation (RAG) Chatbot Effectiveness in an Online Course. *Information Systems Education Journal* v23 (n2) pp 4-13. <https://doi.org/10.62273/ZKLLK5988>
- Leidig, P., Salmela, H., & The Joint ACM/AIS IS2020 Task Force. (2020). IS2020: A Competency Model for Undergraduate Programs in Information Systems, ACM, AIS, <https://doi.org/10.1145/3460863>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Lin, Z., Guan, S., Zhang, W., Zhang, H., Li, Y., & Zhang, H. (2024). Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. *Artificial Intelligence Review*, 57(9), 243. <https://doi.org/10.1007/s10462-024-10896-y>
- Marquardson, J., (2024). Embracing Artificial Intelligence to Improve Self Directed Learning: A Cybersecurity Classroom Study. *Information Systems Education Journal*, 22(1), pp.4-13. <https://doi.org/10.62273/WZBY3952>
- Mew, L., Money, W.H., (2024). Leveraging ChatGPT for Higher Education Course Offerings. *Information Systems Education Journal* 22(5) pp 72-80. <https://doi.org/10.62273/OKLK1331>
- Padovano, A., & Cardamone, M. (2024). Towards human-AI collaboration in the competency-based curriculum development process: The case of industrial engineering and management education. *Computers and Education: Artificial Intelligence*, 7, 100256. <https://doi.org/10.1016/j.caeai.2024.100256>
- Qu, R., Tu, R., & Bao, F. (2024). Is Semantic Chunking Worth the Computational Cost? In *Findings of the Association for Computational Linguistics: NAACL 2025*, pp 2155–2177, Albuquerque, NM. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-naacl.114>
- Sridhar, P., Doyle, A., Agarwal, A., Bogart, C., Savelka, J., & Sakr, M. (2023). Harnessing llms in curricular design: Using gpt-4 to support authoring of learning objectives. *arXiv preprint arXiv:2306.17459*.
- TruLens. (n.d.). TruLens: Evaluating and tracking LLM apps. Retrieved June 19, 2025, from <https://www.trulens.org/>
- vom Brocke, J., Tan, B., Topi, H., Weinmann, M. (2020), AIS Global IS Education Report, The Global Report of the Association for Information Systems on Information Systems Education 2020. 4th Edition. Available at: https://www.researchgate.net/publication/339738830_AIS_Global_IS_Education_Report_201920

APPENDIX A

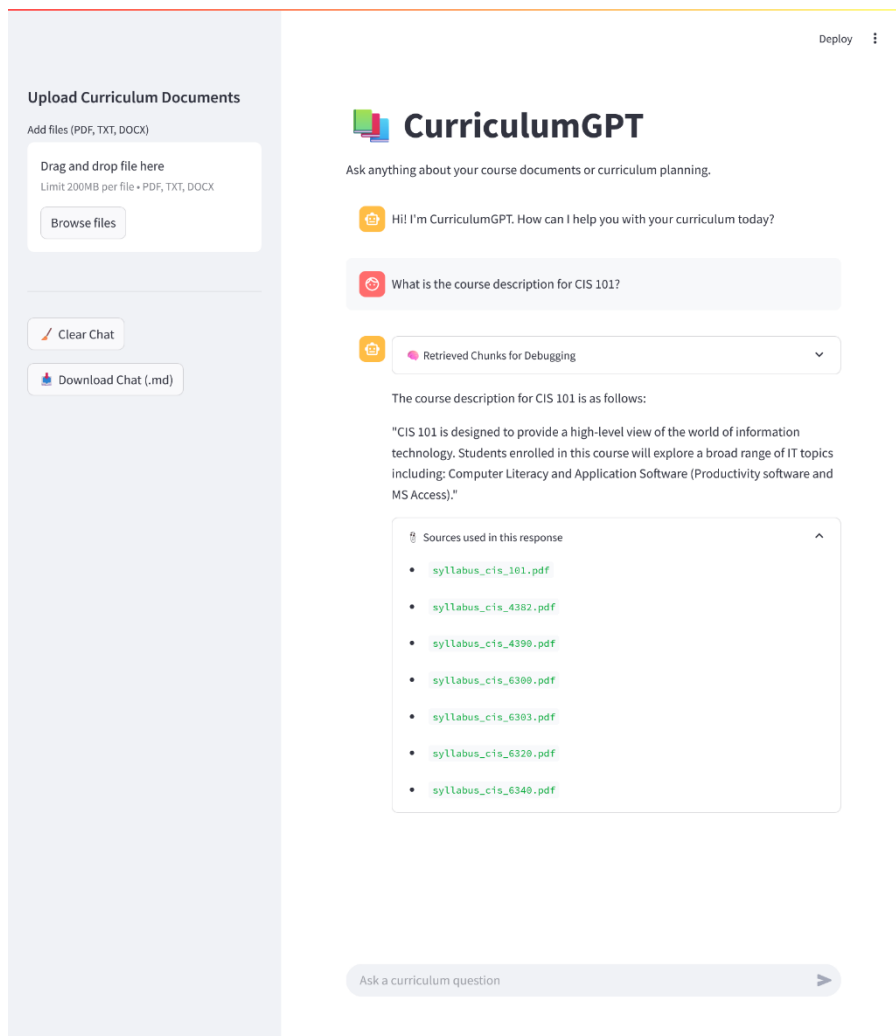


Figure 2: Screenshot of CurriculumGPT

This screenshot shows the CurriculumGPT user interface. Users can interact with it through this web interface by typing questions, seeing responses and citations. Users can also use the sidebar to upload documents, clear, or download chats.

APPENDIX B

CurriculumGPT Evaluation Rubric

1. Faithfulness (Accuracy)
Does the response stay true to the source documents or knowledge base, avoiding hallucinations or invented details?
1 = Incorrect: The response contains significant inaccuracies or information that contradicts the source documents.
2 = Mostly Incorrect: The response contains multiple factual errors, making it unreliable.
3 = Partially Correct: The response is a mix of correct and incorrect information; some key facts are missing or wrong.
4 = Mostly Correct: The response is generally accurate but may contain minor inconsistencies or omissions.
5 = Completely Correct: The response is fully accurate and aligns with the provided source documents.
2. Relevance
Does the response directly address the query and stay on topic?
1 = Irrelevant: The response does not address the query and provides completely unrelated information.
2 = Mostly off-topic: The response touches on the topic but fails to provide a direct or useful answer to the query.
3 = Somewhat relevant: The response contains some relevant information but also includes extraneous details or misses key aspects of the query.
4 = Mostly relevant: The response is on topic and provides a useful answer, but could be more focused.
5 = Fully relevant to the query: The response is directly on topic and provides exactly the information needed to answer the query.
3. Clarity
Is the response well-written, logically structured, and easy to understand?
1 = Very unclear/confusing: The response is difficult to read and understand.
2 = Hard to follow: The response lacks a clear logical flow and may be too generic or off-topic to be truly helpful.
3 = Somewhat clear, needs work: The response is generally understandable but may be too broad.
4 = Clear with good structure: The response is well-written, easy to follow, and appropriately scoped.
5 = Very clear, polished, and coherent: The response is exceptionally well-written and perfectly tailored to the request. It is concise yet comprehensive.
4. Actionability
Would this response be practically helpful for IS educators (e.g., in creating competencies, CLOs, or curriculum analysis)?
1 = Not Actionable: The response is vague, generic, or so poor that it cannot be used for any practical curriculum task.
2 = Minimally Actionable: The response provides very little value and would require extensive human effort to become useful.
3 = Partially Actionable: The response provides a starting point, but significant editing, refinement, or additional research would be needed to make it useful.
4 = Mostly Actionable: The response is a good draft that could be used with minor revisions to complete a curriculum task.
5 = Highly Actionable: The response is a finished or near-finished product that can be used immediately to complete a curriculum task with no or minimal changes.
5. Please comment on any strengths of the AI response and weaknesses or issues: _____

APPENDIX C

Selected CurriculumGPT Queries and Responses

Category 1: System Validation - Basic Retrieval and Fact Recall

Query 1:

"What are the course learning objectives (CLOs) for CIS 3312?"

Response 1:

The Course Learning Objectives (CLOs) for CIS 3312, as outlined in the syllabus, are as follows:

1. **Create data-bound controls for web user interfaces** - This objective focuses on the ability to develop user interfaces that are dynamically linked to data sources.
2. **Separate presentation and business logic in web applications** - This objective emphasizes the importance of distinguishing the different roles in web application architecture, particularly in maintaining cleaner code and enhancing maintainability.
3. **Evaluate business logic with unit tests** - This objective aims to ensure that students are able to apply testing methodologies to the business logic within web applications to verify that the code behaves as expected.

These CLOs align with the overarching goals of critical thinking and functional business analyses, ensuring that students not only learn technical skills but also apply them effectively within the context of web application development (Source: Syllabus for CIS 3312).

Query 2:

"What is a knowledge-skill pair?"

Response 2:

A knowledge-skill pair (K-S pair) is a fundamental concept in competency-based education and curriculum design, integrating both the knowledge component and the corresponding skill level required to successfully perform a specific task or competency.

Definition and Components:

1. **Knowledge Element:** This represents the theoretical or factual information that a learner must understand or know in order to engage in a specific context or task. For example, in computing, a knowledge element could be the "principles of software development."
2. **Skill Level:** This refers to the degree of proficiency or cognitive level associated with applying that knowledge. The skill levels are often classified using Bloom's Taxonomy, which includes levels such as:

- Remember (Level 1)
- Understand (Level 2)
- Apply (Level 3)
- Analyze (Level 4)
- Evaluate (Level 5)
- Create (Level 6).

For instance, a K-S pair could be structured as follows:

- Knowledge Element: "Purpose of readable source code"
- Skill Level: "Understand" (Level 2)

Importance in Curriculum Design:

Knowledge-skill pairs are used to define competencies within educational programs, indicating what learners should know and be able to do in various situations. They enable educators to design courses and assessments that align with specific learning objectives. Each competency includes multiple K-S pairs that encapsulate the knowledge and skills necessary for mastery of that competency.

In essence, K-S pairs help clarify the relationship between what students need to learn (knowledge) and how they need to demonstrate that learning (skill). This structured approach ensures a comprehensive and coherent curriculum that effectively prepares learners for practical applications of their education.

[Reference: IS2020 Curriculum Document; CC2020 Computing Curricula Report]

Query 3:

"What is the course description for CIS 101?"

Response 3:

The course description for CIS 101 is as follows:

"CIS 101 is designed to provide a high-level view of the world of information technology. Students enrolled in this course will explore a broad range of IT topics including: Computer Literacy and Application Software (Productivity software and MS Access)."

Category 2: Comparisons - Syllabus Gap Analysis and Competency Comparison

Query 1:

"What key computing competencies are missing from the CIS 1301 syllabus?"

Response 1:

To identify key computing competencies that may be missing from the CIS 1301 syllabus based on the information provided, we can look at the course description, objectives, and materials in conjunction with contemporary computing education frameworks. Here's a structured analysis:

Overview of CIS 1301 Competencies:

From the syllabus, the primary focus of CIS 1301 is on developing proficiency in Microsoft Office applications (Word, PowerPoint, Excel, and Access) and general computer literacy. The competencies related to personal productivity software aim to build foundational skills that are essential for business and academic activities.

Analysis of Missing Competencies:

1. Critical Thinking and Problem Solving:

- The syllabus emphasizes proficiency in software applications, but there is less attention on developing critical thinking skills, particularly in the context of analyzing and solving problems using IT. According to CC2020 (CC2020, page 42), 81% of employers report a lack of critical thinking and analytical skills among graduates. Integrating objectives that require students to apply software for data analysis or project management could fill this gap.

2. Collaboration and Communication Skills:

- While the course focuses on productivity software, it should also emphasize collaborative tools and skills. Competencies that encourage teamwork (e.g., using tools within Microsoft

Teams or collaborative features in Office 365) are essential in modern workplaces and can help address the communication gap noted by employers (CC2020, page 42).

3. Data Literacy:

- Given the pivotal role of data in contemporary IT environments, a syllabus that encourages understanding of data management principles, along with decisions based on data analysis, is vital. This includes exploring how to structure databases and querying with MS Access, aligning with competencies highlighted in CC2020 regarding data management (CC2020, page 42).

4. Ethics and Professional Practice:

- There is no mention of ethics in technology use, which is increasingly crucial in IT education. Integrating discussions about responsible use of information technology, data privacy, and the implications of technology in society could enhance the curriculum significantly (CC2020, page 46).

5. Emerging Technologies:

- While the course covers basic software applications, it may be beneficial to introduce competencies related to understanding and applying emerging technologies, such as cloud computing, to provide students with a more forward-thinking skill set that aligns with industry needs (CC2020, page 42).

6. Project Management Fundamentals:

- Understanding basic project management concepts, especially as they relate to IT and software projects, would be beneficial. This could be integrated through practical assignments that require planning and executing a project using the tools covered in the syllabus (CC2020, page 46).

7. Practical Application of Skills:

- There is an emphasis on proficiency in software packages, but the syllabus could benefit from more focus on the application of these tools in real-world scenarios. Assessments could include projects that require synthesizing skills across different applications to solve a problem.

Recommendations:

To align CIS 1301 with current competencies in computing, consider the following additions to the syllabus:

- Critical Thinking Projects: Include tasks that require students to synthesize information and make data-driven decisions using the software.
- Collaborative Assignments: Use tools that facilitate teamwork and enhance communication skills.
- Data Literacy Modules: Incorporate database management concepts and data analysis projects using MS Access.
- Ethical Discussions: Integrate conversations about the ethical use of technology.
- Emerging Technology Insights: Familiarize students with trends such as cloud storage and social media analytics.
- Project Management Principles: Introduce basic project management frameworks relevant to IT.

In conclusion, enhancing the CIS 1301 syllabus with these competencies would provide a more comprehensive educational experience, better preparing students for the evolving demands of the information technology landscape.

Query 2:

“Compare the competencies between CIS 4308 and CIS 4350.”

Response 2:

To compare the competencies between CIS 4308 (Introduction to Data Analytics) and CIS 4350 (Data Mining Methods), we first need to examine the primary focus and competencies associated with each course.

CIS 4308: Introduction to Data Analytics

- **Competencies:**
 - Understanding foundational concepts in data analytics, including data collection, preprocessing, and exploratory data analysis.
 - Proficiency in statistical analysis and interpretation of data results.
 - Familiarity with data visualization techniques and tools to communicate findings effectively.
 - Basic skills in programming or software used for analytics (e.g., Python, R, SQL).
 - Application of analytical methods to solve business-related problems.

CIS 4350: Data Mining Methods

- **Competencies:**
 - Advanced understanding of different data mining techniques such as classification, regression, clustering, and association rule mining.
 - Proficiency in using data mining software and frameworks (like R, RapidMiner, or WEKA).
 - Ability to handle large datasets and understand underlying data structures.
 - Skills in developing predictive models and evaluating their performance.
 - Application of ethical considerations and the implications of data mining in business contexts.

Comparison of Competencies

1. Foundational vs. Advanced:

- CIS 4308 focuses on foundational data analytics skills and general data interpretation, making it suitable for beginners or students looking to understand the basics of analytics.
- CIS 4350 delves into advanced data mining techniques, suitable for students eager to specialize in predictive analytics and deeper technical skills.

2. Technical Skills:

- Both courses likely require some technical skills but to different degrees. CIS 4308 may introduce students to analytics languages and tools, while CIS 4350 would require mastery of these tools alongside complex data mining algorithms.

3. Application and Problem Solving:

- In CIS 4308, students will learn to analyze data to inform decisions and provide insights that can guide business strategy.
- CIS 4350 equips students with the skills necessary to create models that predict outcomes based on historical data, ideal for roles in data science and machine learning.

4. **Ethics and Impact:**

- While both courses will likely touch on ethical considerations related to data usage, CIS 4350 may go deeper into the implications of data mining and predictive analytics on privacy and decision-making processes.

In summary, CIS 4308 serves as an entry-level introduction to data analytics, focusing on basic skills and foundational knowledge, while CIS 4350 builds on that foundation by exploring more complex methodologies in data mining, requiring a stronger technical background and analytical capabilities. Each course targets different stages in a student's education and career path within the field of data analytics and information systems.

Category 3: Authoring Competencies and Course Outlines

Query 1:

"Can you create one focused computing competency for my course CIS 3312 that targets a single learning objective?"

Response 1:

Certainly! Here is a focused computing competency for CIS 3312 – Advanced Business Programming:

Computing Competency for CIS 3312 - Advanced Business Programming

Competency Statement:

Students will be able to apply object-oriented programming principles to develop software solutions that address defined business requirements.

Learning Objective:

Apply object-oriented principles (e.g., classes, inheritance, and encapsulation) to design and implement a functional software component that meets specified business requirements.

Query 2:

"Generate a sample syllabus outline for a course that teaches cybersecurity and maps to the CC2020 competencies."

Response 2:

Here is a sample syllabus outline for a course titled "Introduction to Cybersecurity," structured to align with the competencies defined in the CC2020 report. This syllabus is intended to cover the fundamental principles of cybersecurity, emphasizing both technical skills and the necessary communication and strategic thinking required in the field.

Course Title: Introduction to Cybersecurity

Course Description:

This course provides an introduction to the fundamental concepts, principles, and practices of cybersecurity. It covers risk management, the application of security principles, threat identification, and the skills necessary to develop and enforce security policies in various organizational contexts.

Course Objectives:

Upon completion of this course, students will be able to:

1. Evaluate the purpose and function of cybersecurity technologies and their role in organizational risk management.

2. Implement tools and systems to minimize cybersecurity threats.
3. Develop and articulate the policies and procedures necessary for responding to cybersecurity incidents.
4. Analyze risks and recommend mitigation strategies based on best practices in cybersecurity.

Week-by-Week Outline:

Week 1: Introduction to Cybersecurity

- Overview of cybersecurity
- Importance of cybersecurity in today's environment (CSEC2017, IS2020)
- Review of major cyber threats and vulnerabilities

Week 2: Cybersecurity Principles

- Fundamental cybersecurity principles: confidentiality, integrity, and availability
- Understanding risks and their impact on organizations (ITE-CSP-A)

Week 3: Cybersecurity Technologies

- Overview of cybersecurity tools and technologies (e.g., firewalls, intrusion detection systems)
- Evaluation of technologies that reduce risk of data breaches (ITE-CSP-A)

Week 4: Implementing Security Measures

- Practical implementation of security measures
- Hands-on lab: Setting up a firewall and intrusion detection system (ITE-CSP-B)

Week 5: Risk Management Strategies

- Introduction to risk management concepts
- Developing a risk management plan (ITE-CSP-C)

Week 6: Incident Response and Recovery

- Understanding incident response frameworks
- Developing response plans tailored towards high-value systems (ITE-CSP-D)

Week 7: Developing Security Policies

- Crafting and communicating effective security policies
- Case study analysis of organizational failures due to inadequate policies (ITE-GPP-C)

Week 8: Legal and Ethical Considerations in Cybersecurity

- Overview of cybersecurity laws and regulations
- Ethical implications of cybersecurity practices (ITE-GPP-C, IS2020 Competency 8)

Week 9: Social Engineering and Human Factors

- Understanding the role of human behavior in cybersecurity
- Techniques for mitigating human-related vulnerabilities

Week 10: Securing Data and Information

- Best practices for data security: protection of data at rest and in transit (ABET Criteria)
- Hands-on lab: Designing a secure database (ITE-IMA-B)

Week 11: Emerging Threats in Cybersecurity

- Analysis of current and emerging threats
- Discussion of security trends and future challenges in cybersecurity

Week 12: Final Projects and Presentations

- Students present their individual or group projects evaluating a cybersecurity framework of their choice
- Focus on lessons learned and future improvements (CSEC2017, IS2020)

Assessment:

- Participation and attendance
- Weekly quizzes
- Midterm exam
- Final project presentation
- Written report on the final project

Credits: 3 credits

Prerequisites: Introduction to Information Technology or equivalent knowledge.

References:

1. ACM. (2017). Cybersecurity Curricula 2017.
2. CC2020. (2020). Computing Curricula 2020.
3. IS2020. (2020). Information Systems Curriculum Guidelines.

This syllabus outline is structured to provide students with the foundational competencies necessary for success in the cybersecurity field while integrating practical applications and case studies for a comprehensive learning experience. The competencies related to risk management, legal considerations, and technical implementations are emphasized, aligning with the goals outlined in the CC2020 report.

A Comparative Case Study of AI-Enabled and Traditional Experiential Simulations in Operations and Supply Chain Education

Yan Jin
yan.jin@quinnipiac.edu

Ae-Sook Kim
ae-sook.kim@quinnipiac.edu

Tan Gürpınar
tan.gurpinar@quinnipiac.edu

Iddrisu Awudu
iddrisu.awudu@quinnipiac.edu

School of Business
Quinnipiac University
Hamden, CT, USA

Abstract

This study presents a comparative case study examining differences in the pedagogical impact between an AI-enabled negotiation simulation and a traditional non-AI 5S (Sort, Set in Order, Shine, Standardize, Sustain) lean simulation on students' conceptual understanding, reflective thinking, and perceived usefulness. The AI-enabled simulation was implemented in an undergraduate Operations and Supply Chain Management (OSCM) course and designed to engage students across all four stages of Kolb's experiential learning cycle through iterative negotiation rounds with an AI-simulated supplier, while a traditional non-AI 5S simulation served as a conventional experiential benchmark. Using a qualitative-dominant mixed-methods case study approach, data were collected from students' reflection reports, discussion posts, peer responses, and a post-assignment perception survey. Findings indicate that student artifacts associated with the AI-enabled simulation, compared to those from the 5S simulation, reflected higher levels of conceptual understanding and cognitive engagement. However, students' perceived usefulness and value were lower for the AI-enabled activity. Compared with the 5S simulation, the AI-enabled activity required more complex tasks. Although it was associated with deeper understanding and insight, students tended to rate easier tasks as more useful. These mixed results reveal a gap between subjective perceptions and objective learning outcomes and underscore the importance of instructional design that incorporates desirable difficulties. This study contributes to OSCM pedagogy by offering evidence-based insights from a single-course implementation into how technology-enhanced experiential learning may support the development of both technical and durable skills while maintaining theoretical coherence and instructional rigor.

Keywords: Experiential Learning, Kolb's Experiential Learning Theory, AI-Enabled Simulation, Operations and Supply Chain Education, Cognitive Engagement, Generative AI

Recommended Citation: JIN, Y., Kim, A.-S., Gürpınar, T., Awudu, I., (2026). A Comparative Case Study of AI-Enabled and Traditional Experiential Simulations in Operations and Supply Chain Education. *Information Systems Education Journal*, v25(n1) pp 23-36. DOI# <https://doi.org/10.62273/GSPC8748>

A Comparative Case Study of AI-Enabled and Traditional Experiential Simulations in Operations and Supply Chain Education

Yan Jin, Ae-Sook Kim, Tan Gürpınar, Iddrisu Awudu

1. INTRODUCTION

Generative AI tools are widely available, raising a pressing question on how to integrate them into education to make teaching and learning more effective. While educators see the benefits, challenges remain in how to incorporate AI tools into education properly and responsibly to maximize its advantages and minimize potential risks such as plagiarism and misinformation (Kurtz et al., 2024; The Oxford University Press, 2023; Tu, 2024; Wang et al., 2024).

In parallel, the rapid evolution of AI is transforming the labor market, compelling education systems to adapt accordingly. Recent research shows that since 2019, the demand for AI-related skills has been growing at a rate of 21% (Bain, 2025). Equally important, human skills, such as creative thinking, adaptability, and ethical judgment, continue to rise in importance (The World Economic Forum, 2025). Within this evolving landscape, organizations globally face significant challenges in finding talent with the optimal combination of technical and durable human skills (LinkedIn, 2025). These trends raise a critical question: how can we effectively and responsibly integrate AI tools into classroom teaching and learning to equip students with the right mix of those skills?

Prior research on AI integration in education advocates fundamental shifts from traditional passive models toward student-centered approaches that promote active engagement and critical thinking (Kurtz et al., 2024; Muscanell & Robert, 2023; Raptis, 2024). Several studies highlight the role of generative AI as a collaborative learning partner in supporting self-directed learning while requiring students to evaluate information and formulate subsequent prompts (Aladsani, 2025; Raptis, 2024). Rather than viewing AI as a threat, education should strategically utilize AI tools to facilitate interactive and dynamic engagement with educational materials (Association to Advance Collegiate Schools of Business [AACSB], 2023).

Given this emphasis on preparing students for AI-driven work environments, OSCM offers a

compelling context for experiential AI integration. Recent work suggests that AI-supported instructional activities can enhance student engagement and learning when combined with structured guidance and validation tools (Li & Ho, 2024). Shalpegin and Nguyen (2024) argue that incorporating AI education provides opportunities for students to develop skills in technology and automation and offers innovative ways for students to strengthen their AI adaptation and application in the professional world.

In parallel, supply chain management education research argues that students benefit when learning experiences connect conceptual knowledge with real-world application and industry-relevant problem solving (Baveja et al., 2024). Poo and Qi (2023) similarly suggest that AI technologies, such as machine learning algorithms, optimization models, and simulation tools, provide an opportunity for increasing the impact and significance of the quality and effectiveness of laboratory experiments and thereby implicitly impact teaching. They conclude that the AI-powered class sessions empower students with an innate understanding of real-world supply chain challenges. Together, these studies suggest that AI-enabled instructional design may strengthen students' understanding of core OSCM concepts while also supporting the development of practical decision-making skills.

From a corporate perspective, AI and machine learning are increasingly recognized as strategic assets in OSCM, driving improvements across demand forecasting, inventory optimization, and operational execution (Çaylı & Oralhan, 2024; Kumar et al., 2024; Meyer et al., 2021; Mohsen, 2023). However, AI tools are only effective when professionals can interpret and act on their insights (Abhulimen & Ejike, 2024; de la Roche et al., 2024; Forbes, 2025). To address the growing skills gap, organizations are investing in equipping employees to work effectively with AI technologies (Cardon, 2023). Despite these investments, there is still limited research on how AI-based learning activities affect students' level of reflective thinking and understanding of key concepts in OSCM education.

Building on the workforce's need for both AI literacy and domain-specific skills, this study introduces an AI-Enabled Negotiation Simulation (referred to as "AI simulation"), compares it with a non-AI 5S Lean Simulation (referred to as "5S simulation") within an undergraduate OSCM course. The study examines how students' reflective thinking, conceptual understanding, and perceptions differ across these two experiential learning activities. The following research questions guide this inquiry:

- RQ1: In what ways is the AI simulation associated with student engagement in the learning process, particularly in terms of their cognitive effort and reflective thinking?
- RQ2: How is the AI simulation associated with the depth and quality of students' conceptual understanding?
- RQ3: How do students evaluate the AI simulation compared to the 5S simulation?

2. THEORETICAL FRAMEWORK

Kolb's Experiential Learning Theory (ELT) is a useful framework to model a continuous and holistic learning cycle of a student. Kolb (1984) described learning as a dynamic process of grasping experiences and transforming them into knowledge through four stages: concrete experience, reflective observation, abstract conceptualization, and active experimentation. This cyclical process, shown in Figure 1, begins with concrete experience, where learning occurs through doing or observing in real or simulated contexts such as fieldwork, labs, or simulations. In reflective observation, students consciously analyze these experiences, often using brainstorming, discussion, or personal journaling; In abstract conceptualization, they develop theories, models, or hypotheses based on their reflections, drawing meaning from their experiences through logic and analysis; In active experimentation, students test these ideas by planning and applying them in new contexts, which may include further labs, fieldwork, projects, or case studies (Akella, 2010; Bergsteiner et al., 2010; Botelho et al., 2016).

The ELT's dynamic and cyclical nature is well suited to modeling a student's learning of complex, nonlinear phenomena (Botelho et al., 2016). Learning from concrete experience requires active and critical reflection by observing, a student's effort to understand their experience, followed by the conceptualization of a theory or hypothesis by thinking (Botelho et al., 2016). The continuous process will test this theory, which is connected to another concrete

experience (Bell & Bell, 2020; Fewster-Thuente & Batteson, 2018).

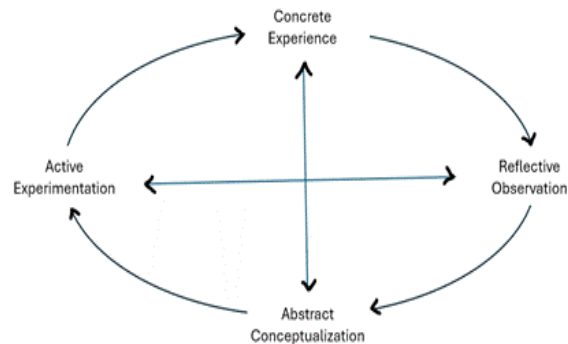


Figure 1: Kolb's Experiential Learning Framework

Note. From *Experiential Learning: Experience as the Source of Learning and Development* (p. 21), by D. A. Kolb, 1984, Prentice-Hall. Copyright 1984 by Prentice-Hall.

Kolb's ELT has been extensively utilized in various fields of education (Bell & Bell, 2020; Botelho et al., 2016; Fewster-Thuente & Batteson, 2018; Turesky & Wood, 2010; Wijnen-Meijer et al., 2022), especially with the application of simulations. This theory also aligns closely with the fundamental transformation of teaching and learning in the age of AI, as both emphasize student-centered approaches that promote self-directed learning while fostering interactive and dynamic cognitive engagement with educational content. In this study, the terms "engagement" and "cognitive engagement" are used interchangeably and refer to the mental effort and strategic thinking students invest in learning tasks (Fredricks, Blumenfeld, & Paris, 2004). The key elements of ELT theory that demonstrate this alignment include the following (Kolb & Kolb, 2009; Morris, 2020):

- Learning is grounded in experience. The AI simulation provides an ideal environment for students to develop a stronger sense of themselves as learners by negotiating procurement contracts with an AI supplier. This activity immerses students through four stages in the ELT model.
- Metacognitive capabilities involve learners' awareness of their learning processes, recognition of challenges, and strategies for improving learning. The AI simulation supports the development of these capabilities by prompting students to identify obstacles and reflect on ways to enhance their learning.
- Learning is a cyclical process, where concrete experiences are continuously enriched through reflection, conceptualization, and active

experimentation. The AI simulation reinforces this cycle, as reaching an agreement typically requires multiple negotiation rounds. Each round builds on the previous one, a spiral of increasingly sophisticated understanding.

- The effectiveness of experiential learning depends on context. The AI simulation provides a rich, discipline-specific environment that reflects real-world negotiation challenges in OSCM. This context informs students' experiences, sharpens their conceptual thinking, and shapes how they perceive and internalize what they learn.

To examine how these experiential learning elements unfold in practice, this study used a qualitative case study approach to compare two simulations (i.e., AI simulation and 5S simulation) and explored how student learning was expressed within the AI simulation context.

3. METHODOLOGY

Research Design

This study adopts a qualitative-dominant mixed-methods case study design. According to Baxter and Jack (2008), the qualitative case study provides guidance for developing clear research questions and the focus of the study, emphasizes theoretically informed propositions to guide data collection and analysis, stresses the importance of "binding the case" for what will and will not be studied, and requires multiple sources of evidence

to support triangulation and strengthen the study's credibility. In this study, qualitative analysis of student reflections and discussion artifacts serves as the primary analytical approach, complemented by quantitative summaries and statistical comparisons to support the interpretation of patterns across simulation activities and students' perceptions.

Combined with Kolb's experiential learning theory, this research design placed students in authentic contexts where they could develop and apply critical thinking, problem-solving, negotiation, and prompt engineering skills. This methodological design ensured that the research questions regarding student learning processes and experiences were explored holistically, in line with the study's aim to understand how AI tools influence applied OSCM education. The study remained bounded within the scope of a single undergraduate OSCM course with clearly defined assignments, aligning with Baxter and Jack's (2008) guidance for maintaining a focused case. Two simulations (AI simulation and 5S simulation) generated rich qualitative data through individual reflection reports, Blackboard discussion posts, and peer replies. These qualitative data were systematically coded and analyzed, with coding frequencies, average scores, and statistical comparisons used to examine differences between the two simulations. These multiple data sources enabled triangulation and supported a credible chain of evidence (Table 1).

RQ	Data Source	Outcome Measures	Analytical Approach
RQ1: Cognitive engagement effort and reflective thinking	Reflection reports Discussion posts/replies	Reflective Insights	Thematic coding; Frequency and average comparisons; ANOVA and T-Test
RQ2: Depth and quality of conceptual understanding	Reflection reports Discussion posts/replies	Conceptual understanding	Thematic coding; Frequency and average comparisons; ANOVA and T-Test
RQ3: Student perceptions	Post-assignment survey	Perceived usefulness and value	Descriptive statistics (distribution of ratings)

Table 1: Research Design Alignment

Participant Recruitment

This study was conducted in a private urban university in the northeastern United States. Participants included 40 students enrolled in an introductory course on OSCM, which is part of the business core curriculum. The sample represented students from both business-related and non-business majors. The participants were sophomores, juniors, and seniors.

Of the 40 students enrolled in the course, 11 were excluded from the analysis due to incomplete submissions, defined as missing one or more

required components (e.g., reflection report, discussion post, or peer reply) across the simulation activities. Only students with complete data across all evaluated components were included to ensure consistency in coding and comparison. A comparison of respondent and non-respondent distributions by major and class indicates noticeable differences (Appendix A), suggesting the possibility of non-response bias. For example, finance majors comprised a larger proportion of respondents but were not represented among non-respondents. In terms of class standing, sophomore students represented

a larger portion of respondents than in the non-respondent group. Due to the small sample size and uneven group distributions, these observations are descriptive and should be interpreted with caution.

Data Collection

The data were collected through two simulation assignments and a post-assignment survey. Both simulations followed an experiential learning approach to actively engage students with OSCM concepts. The key difference was that one simulation integrated AI while the other did not, allowing for a comparison of differences in engagement and perceived learning.

The AI simulation, conducted in week 9 as part of the inventory management chapter, used a generative AI tool (ChatGPT or Copilot) to integrate inventory cost analysis with procurement negotiation techniques. This assignment combined technical skills (e.g., cost modeling) with durable skills (e.g., negotiation, communication, and problem-solving). Students first reviewed how to compare ordering policies and how purchasing terms affect inventory costs, then engaged in multiple negotiation rounds with an AI-simulated supplier. In the negotiation process, students assumed the role of a procurement manager, proposed contract terms, received counteroffers from the AI acting as the supplier, and iteratively refined their negotiation strategies through successive exchanges. Students were also required to document key negotiation rounds, including prompts, counter offers, and final agreements, as part of their reflection. Sample prompts and negotiation guidance (e.g., order quantity, price, and contract terms) were provided to support student interaction with the AI tool. For this simulation, each student submitted an individual reflection report, posted key results and strategies on a Blackboard discussion board, and replied to another student's post.

The 5S simulation, conducted in week 11 as part of the Lean and Six Sigma chapter and hosted on a traditional website, helped students internalize the core Lean principles: Sort, Set in Order, Shine, Standardize, and Sustain (Randhawa & Ahuja, 2017). Through an interactive online game (5S Alphabet Game, 2014), students completed simple sorting and error-finding tasks across six scenarios, experiencing incremental improvements in task efficiency. For this activity, students posted their simulation results and a brief reflection on the discussion board, along with a reply to a classmate's post.

The AI simulation was completed two weeks prior to the 5S simulation and was embedded in a different course chapter, reducing the likelihood of direct knowledge carryover between tasks. However, the study does not include a separate control group or randomized assignment, as all students completed both simulations as part of the course requirements. In addition, no pre-assignment baseline measures were collected. Therefore, the findings should be interpreted as comparative insights within a single-course implementation rather than causal evidence of the effectiveness of AI-enabled versus non-AI simulations.

In addition, a post-assignment survey designed by the instructor and administered at the end of the semester collected student numerical ratings on "perceived usefulness and value" of each simulation assignment. Students evaluated how much they liked each assignment, considering how interesting it was and how much it helped them understand OSCM concepts (Appendix B).

Data Coding Procedure

For qualitative data, such as the reflection reports, discussion posts, and peer replies, a thematic analysis was conducted. This widely used method in qualitative research helps to identify, analyze, and interpret patterns (or "themes") (Braun & Clarke, 2006), making it well-suited for exploring students' learning experiences and perceptions in an educational context.

The coding process began with an initial structure consisting of two broad themes: cognitive and reflective processes and conceptual understanding. Each theme was defined across multiple performance levels, with level 0 representing no meaningful or relevant engagement while level 3 indicating the highest level of involvement. Through careful review of students' reflection reports, discussion posts, and peer replies, relevant keywords were identified, grouped into categories, and aggregated into thematic patterns. These themes were iteratively refined and formalized into a structured set of coding dimensions: *Reflective Insight* and *Conceptual Understanding*. At the same time, the rubric definitions for each performance level were refined to support consistent evaluation. Each student response was evaluated based on the presence, depth, and integration of relevant concepts and reasoning processes. The final coding schema and its detailed rubric are presented in Section 4 and Appendix C.

When responses did not clearly align with a specific performance level within a coding dimension, decisions were made by referencing the rubric definitions to ensure consistent interpretation across submissions. For example, a response that connected negotiation outcomes to inventory cost trade-offs and explained how adjusting order quantity reduced holding costs while maintaining supplier agreement was coded as level 3 in *Conceptual Understanding*. In contrast, a response that merely described the negotiation outcome without explaining the underlying reasoning was coded as level 1. Similarly, for *Reflective Insight*, a response that identified challenges, proposed alternative strategies (e.g., payment terms or credits), and explained how these contributed to the outcome was coded as level 3. In contrast, a response that provided only a general statement of the outcome without explanation or reflection was coded as level 1.

Due to the nature of the data collected, the coding was conducted solely by the course instructor, in accordance with the IRB exemption granted for this study. Because coding was conducted by the course instructor, this analysis emphasizes transparency and procedural consistency rather than coder independence. To support interpretive credibility, a clearly defined rubric with explicit level definitions was applied consistently across artifacts, and coding was conducted in two rounds separated by a one-month interval to enhance intra-rater stability. Multiple data sources (reflection reports, discussion posts, peer replies, and survey data) were used to triangulate patterns across assignments. While these procedures do not substitute for inter-rater reliability, they provide a transparent and theoretically grounded basis for interpretation within a classroom-based qualitative case study.

To complement the qualitative analysis, quantitative summaries were computed from coding scores (0–3) for each dimension. One-way ANOVA was used to compare mean scores across simulation assignments. Given the ordinal nature of the coding scales and the repeated-measures structure of the data, these analyses are used to support descriptive comparison of patterns rather than confirmatory statistical inference.

4. FINDINGS

Description of Participants

Although 40 students were enrolled in the OSCM course, the analysis is based on 29 participants with complete data across all evaluated components. Since the OSCM course is part of the

business core curriculum, 95% of the enrolled students were in business majors. Because of the depth and specialization of the content knowledge required for the course, no freshman students enrolled in the course. The participants' top three majors were finance (32.5%), marketing (22.5%), and applied business (17.5%), which are the common majors among business students in sophomore, junior, and senior classes.

Final Coding Schema

This research focused on comparing the cognitive engagement and depth of understanding demonstrated in reflection reports and two discussion boards. Based on the iterative coding process described in Section 3, a coding schema was developed to capture key dimensions of student learning, including *Reflective Insight* and *Conceptual Understanding*. The dimension *Reflective Insight* was identical for all assignments. To maintain the unique context of the AI Simulation and the 5S simulation, the focus of *Conceptual Understanding* for the AI simulation was on students' understanding and application of inventory management and negotiation principles, such as Economic Order Quantity (i.e., EOQ), inventory costs, negotiation strategies, and contractual terms, while the focus for the 5S simulation was on the key Lean philosophy and 5S principles.

Each dimension within the coding rubric had its own scale, so the same numeric level indicates different evidence depending on the dimension. Level 3 responses in *Conceptual Understanding* demonstrated integrated reasoning across multiple OSCM concepts. For example, students connected negotiation outcomes to inventory cost trade-offs and explained how changes in order quantity, pricing, or contract terms affected cost efficiency. In contrast, lower-level responses described outcomes without linking them to underlying concepts. Similarly, level 3 *Reflective Insight* responses included discussion of challenges, alternative strategies, and how these contributed to the outcome. Lower-level responses were more general, with limited reflection on the process or strategies used. In terms of *Conceptual Understanding*, although the coding rubric specifies different content domains (inventory/negotiation vs. Lean/5S) tailored to each simulation's focus, both were developed to assess the same dimension of *Conceptual Understanding*, which is students' ability to grasp and apply key OSCM principles within the context of each simulation. See Appendix C for the complete coding scale definitions (levels 0–3).

Individual Reflection Reports and Discussion Posts

Each student's reflection reports and discussions (posts and replies) for the AI simulation and discussions (posts and replies) for the 5S simulation were evaluated using a predefined coding schema. No submission received a score of 0 or 1. Therefore, only the frequencies for levels 2 and 3 are shown in Table 2. The AI simulation reflection report received the highest average score for both *Reflective Insight* (2.72) and *Conceptual Understanding* (2.79), followed by the AI simulation discussion (2.66 and 2.62, respectively). The 5S simulation discussion received the lowest scores in both categories (2.31 for each).

Coding Scale	Reflective Insight		
	AI Simulation Reflection	AI Simulation Discussion	5S Simulation Discussion
Level 2	8 (28%)	10 (34%)	16 (55%)
Level 3	21 (72%)	19 (62%)	11 (45%)
Total	29 (100%)	29 (100%)	29 (100%)
Avg	2.72	2.66	2.31
Coding Scale	Conceptual Understanding		
	AI Simulation Reflection	AI Simulation Discussion	5S Simulation Discussion
Level 2	6 (21%)	11 (38%)	16 (55%)
Level 3	23 (79%)	18 (66%)	11 (45%)
Total	29 (100%)	29 (100%)	29 (100%)
Avg	2.79	2.62	2.31

Table 2: The Effects of AI Simulation on Engagement and Understanding

Statistical comparisons were conducted using one-way ANOVA to examine differences in mean scores across simulation activities. The differences in both dimensions between the AI simulation reflection report and other assignments were statistically significant ($p < 0.1$). In addition, the differences between the AI simulation discussion and the 5S simulation discussion were significant for *Reflective Insight* ($p < 0.01$) and *Conceptual Understanding* ($p < 0.01$). A difference of approximately 0.3–0.5 points on a 0–3 scale suggests a shift from moderate to higher levels of engagement and understanding within this sample. As discussed in Section 3, given the ordinal nature of the coding scales and the repeated measures structure of the data, p values are reported as descriptive indicators of relative differences across assignments rather than as definitive evidence of statistical significance.

The post-assignment perception survey

Students' ratings about their perceptions of the value and helpfulness of two simulations were summarized using descriptive statistics, such as the percentage of responses in "Fair" category. Table 3 reports the number of students selecting each rating, with percentages in parentheses indicating the proportion relative to the total sample of 29 students. The 5S simulation discussion that is always used in the class was valued as the most interesting and helpful (93% rated it as good or excellent) compared to the two assignments based on the AI simulation (less than 80% of students rated them as good or excellent). Between the two assignments from the AI simulation, the reflection report (79.3% rated it as good or excellent), which required more effort than the discussion (72.4% rated it as good or excellent), was preferred and perceived as more valuable and useful.

	Very poor + Poor	Fair	Good + Excellent
5S Lean Discussion	1 (3.5%)	1 (3.5%)	27 (93.0%)
AI Simulation Reflection	1 (3.5%)	5 (17.2%)	23 (79.3%)
AI Simulation Discussion	1 (3.5%)	7 (24.1%)	21 (72.4%)

Table 3: Student Ratings for Assignments

5. DISCUSSION

Student Engagement and Cognitive Effort in AI Simulation

Compared to the non-AI 5S Lean Simulation, the AI-Enabled Negotiation Simulation was associated with a notably higher level of cognitive engagement and reflective thinking in the student learning process. These scores suggest that student artifacts from the AI simulation reflected more active processing of experiences, including reflection on challenges and connections to broader learning goals. For example, a student did not reveal the target cost to the supplier in order not to lose control in the negotiation, showing the active cognitive effort and strategic thinking about information asymmetry, a real negotiation tactic.

Student Conceptual Depth and Understanding in AI Simulation

The AI simulation also was associated with a deeper and higher quality level of conceptual understanding within its domain (inventory management and negotiation principles) compared to the 5S simulation within its domain (Lean/5S principles). This pattern indicates that students' work in the AI simulation more

frequently reflected integrated and comprehensive use of key OSCM concepts. For example, many students reflected that the total inventory cost was influenced by multiple factors, and thus, a low-cost deal isn't always the best deal. For example, students described how adjusting pricing, order volume, and payment structures could significantly influence total cost outcomes. Others emphasized the importance of leveraging supplier relationships to create flexibility in negotiation. These reflections on strategic decisions and holistic cost analysis illustrate how the interconnected nature of procurement negotiation and inventory management decisions was articulated in their written work.

Students' Perceptions: AI Simulation vs. 5S Simulation

Students rated the AI simulation lower than the 5S simulation in terms of "perceived usefulness and value," as reflected in the post-assignment survey. The 5S Simulation has a more mechanical nature and has been used for many years in the course and thus has well-refined instructions. In contrast, the AI simulation is a new and more complex assignment. This unfamiliarity, combined with the higher cognitive demands and less structured nature of the AI interaction, may have contributed to the lower perceived usefulness despite stronger learning outcomes.

This contrast illustrates a typical example of the "Paradox of Actual Learning" (Soderstrom & Bjork, 2015). As Soderstrom and Bjork (2015) pointed out, learners tend to rate easier, more familiar tasks as more useful, even when more challenging activities yield better understanding and insight. In other words, students' perceptions of learning effectiveness or value often diverge from their actual learning outcomes. In this study, although students perceived the 5S simulation as more useful and valuable, the AI simulation demands more sustained thinking and reflective judgment. This pattern aligns with the "desired difficulties" concept (Bjork, 1994; McDaniel & Butler, 2011), which describes how the meaningful challenges can make the initial learning harder but ultimately improve understanding and retention with deeper cognitive engagement. This divergence between perceived value and actual learning underscores the importance of interpreting student satisfaction surveys with caution, especially when evaluating innovative, cognitively demanding activities, such as the AI simulation.

From a pedagogical perspective, this suggests that instructors integrating AI-enabled

simulations should provide clear guidance (e.g., step-by-step negotiation guidelines), scaffolding (e.g., sample prompts, staged reflection questions), and expectation setting to help students navigate initial difficulty. Structured reflection questions and transparent grading criteria may also help students better recognize the learning value of more cognitively demanding activities.

The Alignment with Kolb's Model

Both AI simulation and 5S simulation align well with Kolb's experiential learning theory, particularly the stages of concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb, 1984). From Appendix D, the AI simulation demonstrates deeper iterative engagement through repeated negotiation rounds, consistent with findings from students' reflections, discussion posts, and peer replies. In the concrete experience stage, students received an inventory management scenario with key variables, such as demand, pricing, and cost parameters. They were given practical preparatory materials, including what-if analyses, negotiation tips, and sample AI prompts, before entering a dynamic negotiation with an AI supplier. During the negotiation process, students interacted with the AI supplier for several rounds and experienced trial and error, strategic adjustment, and decision-making under constraints, all of which are core aspects of experiential learning.

Students engaged in the reflective observation stage through each round of negotiation by analyzing both their achievements and setbacks in the simulation. Many students referenced moments of challenge (e.g., rejected offers), which led them to rethink their original strategies. These reflections often progressed into abstract conceptualization, where students formulated general principles about negotiation and inventory cost analysis (e.g., the trade-offs between multiple cost drivers). In the final stage, active experimentation, students applied these insights in subsequent negotiation rounds, adjusting their counter offers to reflect a more holistic and strategic perspective (e.g., requesting a credit reimbursement to offset the high price or offering longer contract terms in exchange for a lower price).

Overall, the AI simulation was designed to align with all four stages of Kolb's (1984) model, and student reflections and discussion artifacts suggest engagement across these stages. Importantly, the simulation's iterative design allows students to cycle through Kolb's model

multiple times. Since the cost target cannot be reached in a single negotiation round, each new contract outcome becomes part of the next concrete experience, initiating a new round of learning. This spiral learning process deepens student engagement and reinforces the connection between all four stages of Kolb's model, which is less evident in 5S simulation. As a result, the AI simulation aligns with Kolb's (1984) framework and illustrates how structured iteration may extend experiential learning cycles within an applied OSCM context.

6. CONCLUSION AND FUTURE WORK

The AI Simulation was associated with deeper conceptual understanding and more reflective insights than the 5S Lean Simulation, despite the AI-enabled assignments receiving lower student ratings for perceived value and usefulness. The findings align with prior study (Poo & Qi, 2023) and Kolb's experiential learning cycle (Kolb, 1984), illustrating how AI-enabled activities may be associated with higher levels of cognitive effort and critical thinking within a single-course implementation. This highlights the pedagogical value of designing "desired difficulties" in experiential learning to challenge students meaningfully (Bjork, 1994; McDaniel & Butler, 2011). The study offers insights into how Kolb's experiential learning cycle may operate in a more iterative, spiral-like process through multiple negotiation rounds with iterative reflection (Appendix E). In addition, this research advances OSCM teaching practices by demonstrating how AI can be responsibly integrated to build both technical (e.g., order quantity calculation) and durable human skills (e.g., negotiation), preparing students for complex OSCM roles that demand agility, data-driven decision-making, and negotiation competence.

Practically, the gap between AI tools' availability and the workforce's ability to use them effectively remains a critical challenge for businesses. Patterns observed in student engagement, reflective thinking, and conceptual understanding suggest that educational programs integrating AI literacy and experiential learning can help address this skills gap. By preparing future supply chain professionals with practical experience in AI-enabled decision-making, organizations can build a workforce better equipped to leverage AI capabilities for strategic advantage. Practitioners can apply these insights by collaborating with educational partners to design hands-on, AI-enabled simulations that encourage learners to engage deeply, reflect critically, and develop problem-solving strategies relevant to fast-

evolving supply chain environments where agility and data-driven strategies are increasingly important.

While this study was conducted as an exploration of a single course implementation, the findings provide a foundation for the future iterative refinement of the AI Simulation and the development of more structured evaluation tools. Future research may extend this work through larger samples, independent coding with a second coder, and longitudinal analysis, to improve the credibility and robustness of findings over time. These adjustments would enable deeper investigation of the pedagogical impact of AI integration on student learning outcomes in OSCM education.

7. REFERENCES

- 5S Alphabet Game (2014, January 31). *Playing the 5S Game Overview - 5S Educational Tool*. <http://5salphabetgame.blogspot.com/2014/01/playing-5s-game-overview.html>
- Abhulimen, A. O., & Ejike, O. G. (2024). Solving supply chain management issues with AI and Big Data analytics for future operational efficiency. *Computer Science & IT Research Journal*, 5(8), 1780-1805. <https://doi.org/10.51594/csitrj.v5i8.1396>
- Akella, D. (2010). Learning together: Kolb's experiential theory and its application. *Journal of Management and Organization*, 16(1), 100-112. <https://doi.org/10.5172/jmo.16.1.100>
- Aladsani, H. K. (2025). Developing postgraduate students' competencies in generative artificial intelligence for ethical integration into academic practices: A participatory action research. *Interactive Learning Environments*, 1-19. <https://doi.org/10.1080/10494820.2025.2487517>
- Association to Advance Collegiate Schools of Business Thought Leadership (AACSB) (2023, November 29). Business school innovation in the age of generative AI. <https://www.aacsb.edu/insights/briefings/business-school-innovation-in-the-age-of-generative-ai>
- Bain (2025). AI: The ambitions are bold, but the talent is scarce. <https://www.bain.com/insights/ai-the-ambitions-are-bold-but-the-talent-is-scarce-snap-chart/>
- Baveja, A., Greeley, L., & McLaury, W. (2024). A participatory framework for bridging the conceptual and talent gaps in supply chain management education. *Decision Sciences*

- Journal of Innovative Education*. 22, 33–49.
<https://doi.org/10.1111/dsji.12304>
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, 13(4), 544-559.
<https://doi.org/10.46743/2160-3715/2008.1573>
- Bell, R., & Bell, H. (2020). Applying educational theory to develop a framework to support the delivery of experiential entrepreneurship education. *Journal of Small Business and Enterprise Development*, 27(6), 987-1004.
<https://doi.org/10.1108/JSBED-01-2020-0012>
- Bergsteiner, H., Avery, G. C., & Neumann, R. (2010). Kolb's experiential learning model: Critique from a modelling perspective. *Studies in Continuing Education*, 32(1), 29–46.
<https://doi.org/10.1080/01580370903534355>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
- Botelho, W. T., Marietto, M. Das G. B., Ferreira, J. C. da M., & Pimentel, E. P. (2016). Kolb's experiential learning theory and Belhot's learning cycle guiding the use of computer simulation in engineering education: A pedagogical proposal to shift toward an experiential pedagogy. *Computer Applications in Engineering Education*, 24(1), 79-88. <https://doi.org/10.1002/cae.21674>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp0630a>
- Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2023). Competencies needed by business professionals in the AI age: Character and communication lead the way. *Business and Professional Comm. Quarterly*, 87(2), 223-246.
<https://doi.org/10.1177/23294906231208166>
- Çaylı, O., & Oralhan, Z. (2024). Artificial Intelligence-driven inventory management: Optimizing stock levels and reducing costs through advanced machine learning techniques. *The European Journal of Research and Development*, 4(4), 427–439.
<https://doi.org/10.56038/ejrnd.v4i4.615>
- de la Roche, M., Voloder, E., Banerjee, A., et al. (2024). Report on Artificial Intelligence and Blockchain convergences. INATBA Report.
<https://dx.doi.org/10.2139/ssrn.5023415>
- Fewster-Thuente, L., & Batterson, T. J. (2018). Kolb's Experiential Learning Theory as a theoretical underpinning for interprofessional education. *Journal of Allied Health*, 47(1), 3-8. <https://www.jstor.org/stable/48722018>
- Forbes (2025). The AI skills gap is slowing down supply chains.
<https://www.forbes.com/sites/kolawolesamu/eladebayo/2025/04/25/the-ai-skills-gap-is-slowing-down-supply-chains/>
- Kolb, A. Y., & Kolb, D. A. (2009). The learning way: Meta-cognitive aspects of experiential learning. *Simulation & Gaming*, 40(3), 297-327.
<https://doi.org/10.1177/1046878108325713>
- Kolb, D. A. (1984). *Experiential Learning: Experience as the source of learning and development*. Prentice-Hall, Upper Saddle River, NJ.
- Kumar, P., Amosu, O. R., & Ogunsuji, Y. M. (2024). Revolutionizing sourcing with AI: Harnessing technology for unprecedented efficiency and cost savings. *World Journal of Advanced Research and Reviews*, 23(2), 925–938.
<https://doi.org/10.30574/wjarr.2024.23.2.2357>
- Kurtz, G., Amzalag, M., Shaked, N., Zaguri, Y., Kohen-Vacs, D., Gal, E., Zailer, G., & Barak-Medina, E. (2024). Strategies for integrating generative AI into higher education: Navigating challenges and leveraging opportunities. *Education Sciences*, 14(5), 503-513.
<https://doi.org/10.3390/educsci14050503>
- Li, Y., & Ho, J. (2024). Generative AI for Operations Management Education and Learning: A Critical Assessment and Analysis. *International Journal of Services and Standards*, 14 (4),
<http://dx.doi.org/10.1504/IJSS.2024.10072276>
- LinkedIn (2025, January). Work Change Report: AI is coming to work.
<https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/Work-Change-Report.pdf>
- McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175–198). Psychology Press.
- Meyer, D., Brüggelolte, M., Gürpınar, T., & Henke, M. (2021). Assistance for the implementation of AI in procurement - An

- analysis of maturity models. In *Proceedings of the 30th IPSERA Conference*.
- Mohsen, B. (2023) Impact of Artificial Intelligence on Supply Chain Management Performance. *Journal of Service Science and Management*, 16, 44-58. <https://doi.org/10.4236/jssm.2023.161004>
- Morris, T. H. (2020). Experiential learning – a systematic review and revision of Kolb’s model. *Interactive Learning Environments*, 28(8), 1064–1077. <https://doi.org/10.1080/10494820.2019.1570279>
- Muscanello, N., & Robert, J. (2023, February 14). EDUCAUSE QuickPoll results: Did ChatGPT write this report? EDUCAUSE. <https://er.educause.edu/articles/2023/2/educause-quickpoll-results-did-chatgpt-write-this-report>
- Poo, M. C. P., & Qi, B. (2023, December). Review on the Applications of AI in Laboratory Experiments in Supply Chain Management Education. In *2023 4th International Conference on Computers and Artificial Intelligence Technology (CAIT)* (pp. 110-114). IEEE.
- Randhawa, J. S., & Ahuja, I. S. (2017), Evaluating impact of 5S implementation on business performance, *International Journal of Productivity and Performance Management*, 66(7), 948-978. <https://doi.org/10.1108/IJPPM-08-2016-0154>
- Raptis, P. R. (2024). Using generative AI to enhance teaching and learning. *Ohio Communication Journal*, 62, 53–59.
- Shalpegin, T., & Nguyen, A. (2024). AI in Supply Chain Education. In X. O’Dea and D.T.K. Ng (Ed.), *Effective Practices in AI Literacy Education: Case Studies and Reflections* (pp. 71-79). Emerald Publishing Limited <https://doi.org/10.1108/978-1-83608-852-320241008>
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199. <https://doi.org/10.1177/1745691615569000>
- The Oxford University Press (2023, October 18). AI in education: where we are and what happens next. <https://corp.oup.com/feature/ai-in-education-where-we-are-and-what-happens-next/#a-global-picture>
- The World Economic Forum (2025, January). Future of jobs report 2025. https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf
- Tu, H. (2024). Responsible use of AI in college teaching and learning: A literature review. *Journal on Excellence in College Teaching*, 35(4), 41-50.
- Turesky E. F., & Wood, D. R. (2010). Kolb's experiential learning as a critical frame for reflective practice. *Academic Leadership*, 8(3), 1-21. <https://doi.org/10.58809/HFJJ1868>
- Wang, N., Wang, X., & Su, Y. S. (2024). Critical analysis of the technological affordances, challenges and future directions of Generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 44(1), 139–155. <https://doi.org/10.1080/02188791.2024.2305156>
- Wijnen-Meijer, M., Brandhuber, T., Schneider, A., Berberat, P. O. (2022). Implementing Kolb’s Experiential Learning Cycle by Linking Real Experience, Case-Based Discussion and Simulation. *Journal of Medical Education and Curricular Development*, 9, 1-5. <https://doi.org/10.1177/2382120522109151>

Appendix A

Distribution of Respondents and Non-Respondents by Student Majors and Class Standing

	Class Demographics N (%)	Study Participants N (%)	Study Nonparticipants N (%)
Major			
Finance	13 (32.5%)	13 (44.8%)	0 (0%)
Marketing	9 (22.5%)	8 (27.6%)	1 (9.1%)
Applied Business	7 (17.5%)	3 (10.3%)	4 (36.4%)
Accounting	3 (7.5%)	3 (10.3%)	0 (0%)
Entrepreneurship	2 (5%)	0 (0%)	2 (18.2%)
Bio Marketing	1 (2.5%)	1 (3.4%)	0 (0%)
Business Analytics	1 (2.5%)	0 (0%)	1 (9.1%)
Sports Management	1 (2.5%)	0 (0%)	1 (9.1%)
Business Undecided	1 (2.5%)	0 (0%)	1 (9.1%)
Non-Business Major	2 (5%)	1 (3.4%)	1 (9.1%)
Total	40 (100%)	29 (100%)	11 (100%)
Class Standing			
Sophomore	24 (60%)	21 (72%)	3 (27.3%)
Junior	12 (30%)	7 (24.1%)	5 (45.5%)
Senior	4 (10%)	1 (3.4%)	3 (27.3%)
Total	40 (100%)	29 (100%)	11 (100%)

Appendix B:

Post-Assignment Survey Questions

Please rate how much you liked the **Discussion #3: Simulation results discussion (procurement negotiation simulation)** overall, considering how interesting it was and how much it helped you understand the operations and supply chain management.

1. Very poor
2. Poor
3. Fair
4. Good
5. Excellent

Please rate how much you liked the **Discussion #4: 5S game and lean** overall, considering how interesting it was and how much it helped you understand the operations and supply chain management.

1. Very poor
2. Poor
3. Fair
4. Good
5. Excellent

Appendix C

Coding Schema for Conceptual Understanding and Reflective Insight

Theme	Depth Levels			
Conceptual Understanding (AI-Enabled Negotiation Simulation) Demonstrates understanding and varying levels of application of key inventory management and negotiation principles.	0 = Absent No relevant concepts are identified or applied.	1 = Surface-level Provides a basic or incomplete response (e.g., mentions cost analysis, negotiation approach, or contract terms) without explanation or clear reasoning.	2 = Applied Demonstrates a clear understanding and partial application of inventory management and negotiation principles (e.g., explains cost analysis or negotiation approach with some reasoning).	3 = Integrated and insightful Demonstrates a comprehensive and well-reasoned application of multiple concepts (e.g., full cost analysis, strategic negotiation with clear rationale, and justified contract terms with impact).
Conceptual Understanding (Non-AI 5S Lean Simulation) Demonstrates understanding and varying levels of application of key lean philosophy and 5S principles.	0 = Absent No relevant Lean or 5S concepts are identified or applied.	1 = Surface-level Provides an incomplete or basic response with limited explanation of Lean philosophy or 5S principles.	2 = Applied Demonstrates a clear understanding and partial application of Lean philosophy and 5S principles (e.g., explains how specific steps improve organization or efficiency with some reasoning).	3 = Integrated and insightful Demonstrates a comprehensive and well-reasoned application of multiple Lean and 5S principles (e.g., integrates several steps to explain process improvements with clear rationale and impact).
Reflective Insight (AI-Enabled Negotiation Simulation and non-AI 5S Lean Simulation) Reflects on learning, challenges, and implications for professional practice	0 = Absent No meaningful reflection on the experience is provided.	1 = General Provides a vague reflection on the experience, with limited detail about what was learned and little discussion of challenges.	2 = Specific Provides a specific reflection on the experience, describing particular challenges or strategies and offering some explanation of what was learned.	3 = Deep and personal Provides a deep and personal reflection on the experience, including detailed discussion of challenges and strategies, and clear insights into learning, growth, or future application.

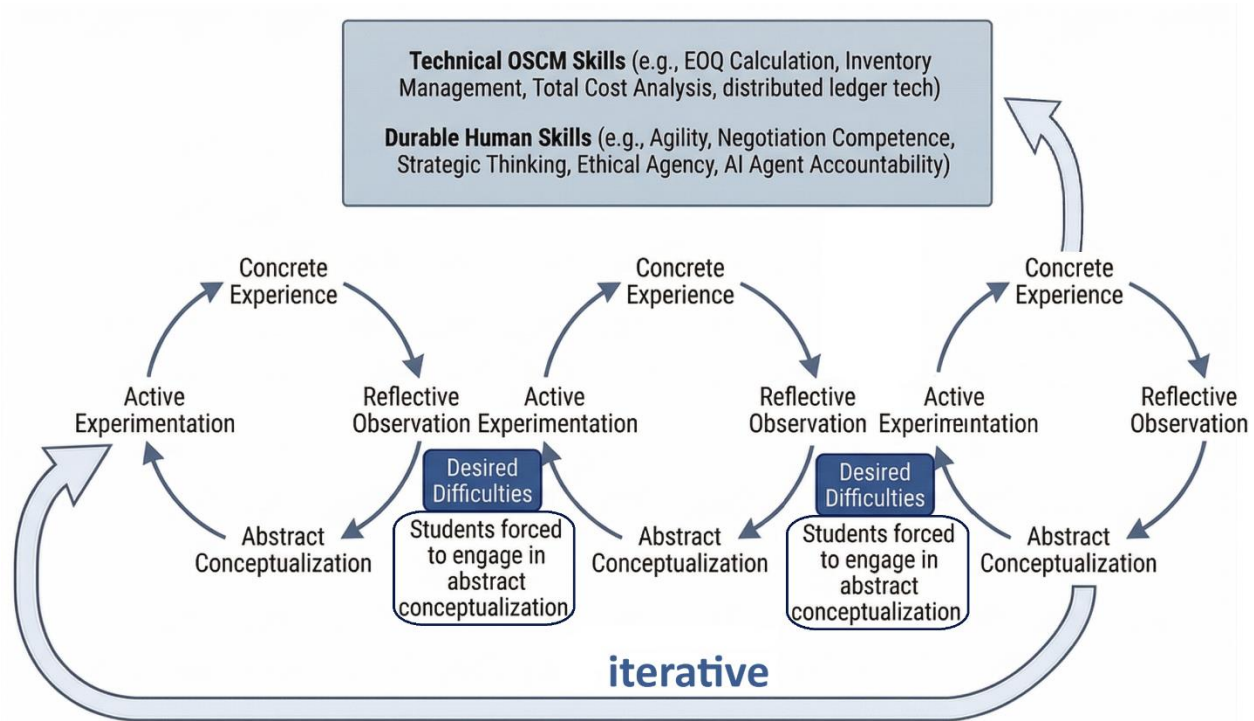
Appendix D

Alignment of Simulations with Kolb's Experiential Learning Cycle

Kolb's Stage	AI-Enabled Negotiation Simulation	5S Lean Simulation
Concrete Experience	Work with inventory management scenarios (demand, pricing, cost variables); engage in multi-round negotiations with an AI supplier using preparatory tools (what-if analyses, negotiation tips, AI prompts).	Perform sorting and error-finding tasks in interactive 5S game scenarios; experience incremental task improvements.
Reflective Observation	After each negotiation round, analyze outcomes, setbacks, and rejected offers; reflect in reports and discussion posts.	Reflect on efficiency gains and mistakes; share brief insights in discussion posts.
Abstract Conceptualization	Formulate general principles about negotiation, trade-offs, and inventory cost drivers; connect to OSCM theory.	Generalize Lean/5S principles; connect task outcomes to workplace organization concepts.
Active Experimentation	Apply revised strategies in subsequent negotiation rounds (e.g., adjust counteroffers, modify contract terms); cycle iteratively through Kolb's stages.	Apply 5S principles in later scenarios or envision applications to daily/work contexts; typically completes a single cycle.

Appendix E

Theoretical Contribution



Personality, Organizational Environment, and the Use of Generative AI: A Conceptual Framework for Understanding Human-AI Interaction

David Firth

David.Firth@umontana.edu
Management Information Systems Department
College of Business, University of Montana
Missoula, MT, 59812, USA

Simon Laub

Sila@eaaa.dk
Business Academy Aarhus,
Aarhus, Denmark

Elizabeth O. Kohl

Elizabeth.Kohl@umontana.edu
Accounting and Finance Department
College of Business, University of Montana
Missoula, MT, 59812, USA

Fang Chen

Fang1.Chen@umontana.edu
Management Information Systems Department
College of Business, University of Montana
Missoula, MT, 59812, USA

Abstract

As generative AI tools become increasingly embedded in education, work, and creative practice, understanding the psychological factors that shape their use is essential. This paper proposes a novel framework that integrates the Big Five personality traits (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) with three key organizational environment task-based motivations, (1) "It's important not to fail", (2) "The output can easily be seen to be correct", and (3) "New ideas are valued", which we use to explain how individuals interact with generative AI systems such as ChatGPT, DALL·E, and Grammarly. While prior research has explored demographic or task-type predictors of AI use, this paper proposes consideration of the individual user's personality and motivation as core variables. We argue that personality traits moderate the likelihood, style, and depth of AI engagement for different types of organizational environment motivation. For example, an individual high in Openness would be drawn to generative AI when novelty is valued by the organization, but less motivated by situations where risk-avoidance is valued. Conversely, an individual high in Neuroticism would engage heavily with AI when failure must be avoided, but is less comfortable with creative ambiguity. Using Weick's (1995) nascent theory approach we have developed a matrix of predicted AI use patterns across combinations of personality profiles and organizational environment motivations. This framework contributes to a more nuanced understanding of human-AI interaction by accounting for individual differences in personality, and has practical implications for the ethical, effective, and inclusive deployment of generative AI technologies.

Keywords: Generative AI, human-AI interaction, task type, personality traits, Big Five personality traits

Recommended Citation: Firth, D., Laub, S., Kohl, E., Chen, F., (2026). Personality, Extrinsic Task Motivation, and the Use of Generative AI: A Framework for Understanding Human-AI Interaction. *Information Systems Education Journal*, v25(n1) pp 37-54. DOI# <https://doi.org/10.62273/FRDJ9146>

Personality, Organizational Environment, and the Use of Generative AI: A Conceptual Framework for Understanding Human-AI Interaction

David Firth, Simon Laub, Fang Chen and Elizabeth Kohl

1. INTRODUCTION

Generative AI tools such as ChatGPT, DALL·E, and Grammarly are becoming extensively used in educational, business, and personal contexts. Humans, as users of Generative AI, have varying personalities and traits, and these characteristics impact how individuals utilize new technologies (Joshi et al., 2023; Labti & Bourhim, 2022). In addition, organizations in which Generative AI is used, including businesses and education, vary with respect to their environment - goals, motivations, internal processes, culture, management approach, and how they engage with the world. Considering the system of human-AI interaction and organizational environment, the research question for this paper is: how do the different personalities and traits of people using Generative AI interact with the ways that organizations want to complete the tasks they are most focused on? This paper uses a grounded theory approach to provide a conceptual prediction framework for the interaction between an individual's personality as described by the Big Five personality traits (see Appendix B), and the task-based motivations of the organization they are in. We use exploratory data to show how the framework might work. The theory presented is at the individual level of analysis as each person has their own personality.

2. NASCENT THEORY DEVELOPMENT

A seminal article by Karl Weick (1995) in *Administrative Science Quarterly* introduced the idea of Nascent Theory to describe early-stage theoretical work that is exploratory, generative, and conceptually creative, has not yet been empirically validated, but is nonetheless valuable for shaping future research. Nascent Theory focuses on novel ideas rather than confirming existing ones. Nascent theory offers new constructs, mechanisms, or relationships that haven't been formally theorized before. The theory is usually supported by logical arguments, illustrative examples, or conceptual reasoning, not empirical testing (yet). The idea is to stimulate scholarly conversation or open a research agenda. Weick argues that the act of constructing ideas, "theorizing", is as important as testing hypotheses. Others, such as Gregor

(2006) in *MIS Quarterly*, have referenced Weick's "What Theory is Not, Theorizing Is" as foundational to understanding theory as an evolving conceptual process, as we do here for this framework.

Existing research has shown that there are various ways that a business can organize, be managed, and operate that can shape how their employees' work. Although these various ways of being organized and managed are extensive, for this framework development, we use Peter Drucker's *The Theory of a Business* (1994) and propose using just three basic organizational emphases: 1) *It's important not to fail*, 2) *Output correctness*, and 3) *New ideas are valued*.

It is important to note that organizational environments are externalities to the user of Generative AI, features of the organization or situation, not of the person. As this paper will explain, this distinction is crucial for understanding how organizational environments interact with employee behavior, including the adoption and use of Generative AI tools.

Personality traits play a critical role in shaping how individuals behave, make decisions, and interact with their environment, including in organizational, technological, and learning contexts (e.g., Chau, 2024; Barczyk & Duncan, 2017; Hignite et al., 1999; Barnett et al., 2015; Liu & Chen, 2025; Reynolds et al., 2017; Woszczyński et al., 2005). Among the most widely accepted models in psychology is the Five-Factor Model (FFM), often referred to as the Big Five personality traits. This framework conceptualizes personality along five broad dimensions: openness to experience (openness), conscientiousness, extraversion, agreeableness, and neuroticism (Costa & McCrae, 1992; John et al., 2008). Each trait captures a spectrum of behavioral tendencies that are relatively stable across time and contexts, making the Big Five particularly useful for research into enduring psychological differences across populations. As a result, the model has been extensively applied in fields ranging from organizational behavior and education to technology adoption and consumer behavior (McCrae & Costa, 1999; Judge et al., 2002).

In recent years, the Big Five framework has gained renewed attention for its explanatory power in studies of individual engagement with new technologies, especially under conditions of uncertainty or innovation (e.g., Joshi et al., 2023; Prayoga & Abraham, 2016; Tripathi et al., 2022). For example, research has shown that individuals high in openness to experience (openness) are more inclined to explore novel tools and adapt to emerging technologies, while those high in conscientiousness may be more structured and deliberate in their technology use (Devaraj et al., 2008; Svendsen et al., 2013; Labti & Bourhim, 2022). Similarly, neuroticism has been associated with anxiety about new systems, while extraversion and agreeableness often influence collaborative and communicative behaviors in digital environments.

Next, we discuss the three types of organizations. Following that we discuss the Big Five personality traits in more detail.

3. ORGANIZATIONAL ENVIRONMENTS

As we build our basic framework, in order to be workable, we decided to limit the types of organizations in the framework to just three. We used Drucker's *The Theory of Business* (1994) to select three organization types that are largely mutually exclusive: "It is important not to fail", "New ideas are valued", and "Output correctness". We recognize that other choices could be made, and that this is indeed a gross simplification of the myriad types of business, but it does provide a starting point.

The *Theory of Business* (Drucker, 1994) is built around three core assumptions that define how an organization operates: 1) Assumptions about the environment – what the organization believes about the world it operates in, 2) Assumptions about the mission – what the organization believes it is there to do, and 3) Assumptions about core competencies – what the organization believes it does well. The three types of organization that we use for our framework have largely different assumptions and so represent a broad range of business types, as shown by the examples below.

1. "It is important not to fail"

The "It is important not to fail" organization has a low risk-tolerance. This shows up in the three core assumptions as follows:

Environment Assumption: The external landscape is high-risk or strictly regulated. Here, failure

carries severe legal, safety, or reputational consequences.

Mission Assumption: The organization's purpose is to deliver absolute reliability, consistency, and safety.

Core Competency Assumption: The organization's primary strength lies in rigorous risk management, compliance, and process control.

Examples: IT functions, aerospace, healthcare, nuclear energy, financial services, and auditing.

Our own experience and observation is that IT organizations are typically of the type that is "important not to fail." It is important that the IT function delivers sufficient processing power, reliability, and accessibility. Though that does not provide any strategic advantage, if it is done poorly it can have a significant negative effect, and hence the need to avoid failure (Carr, 2003).

2. "New ideas are valued"

The "New ideas are valued" organization typically adopts assumptions centered on growth, agility, and market creation rather than just preservation. It is innovation driven, and this shows up in the three core assumptions as follows:

Environment Assumption: The external world is dynamic, competitive, and rewards innovation.

Mission Assumption: The organization exists to create value through innovation and stay ahead of change.

Core Competency Assumption: The organization excels at creativity, experimentation, and rapid adaptation.

Examples: Tech startups, research and development labs, design firms.

3. "Output correctness is most important"

The "Output correctness is most important" organization focus shifts from the low risk tolerance (or the innovation driven) to the absolute accuracy and integrity of the final product. In these "output correctness" organizations, what is important is the verifiable quality of the output. The organization is accuracy driven, and this shows up in the three core assumptions as follows:

Drucker's Assumptions	Organization Type (Environment)		
	Important Not to Fail	New Ideas Valued	Output Correctness
Environment	Regulated & High-Risk	Dynamic & Emerging	Evidence-based & Exacting
Mission	Reliability & Safety	Growth & Discovery	Precision & Integrity
Core Competency	Compliance & Control	Agility & Creativity	Validation & Accuracy

Table 1: Three Organizational Types as aligned with Drucker's Three Core Assumptions Defining How an Organization Operates

Environment Assumption: The external world demands accuracy, precision, and technical excellence.

Mission Assumption: The organization exists to deliver correct, high-quality outputs that meet exacting standards.

Core Competency Assumption: The organization excels at technical expertise, quality assurance, and systematic execution.

Examples: Engineering firms, software quality assurance teams, scientific publishing.

Table 1 (above) summarizes our three types of organizational environments (externalities) along Drucker's (1994) three core assumptions that define how an organization operates. In order to be practical in our framework development, we have limited the number of organizational environments to the three listed here. We are aware that many other types of organizations exist, and save for future research expanding the types of organizations included in our framework.

These three organizational types function as environmental antecedents that shape how tasks are motivated and sustained. In "No Fail" organizations, the focus is on compliance and security. The task motivation is primarily preventative. The main motivational driver in environments where it is important not to fail is fear of negative consequences (legal action, loss of accreditation, or safety incidents). In "New Ideas" environments, such innovation-driven organizations, the culture focuses on growth and recognition. The main motivational driver in an environment where new ideas are valued is seeking rewards, such as career advancement, social recognition within the firm, or "first-mover" status. In "Correctness of Output" organizations the motivation is focused on competence and credibility. The culture is motivated towards precision. The main motivational driver in an

output correctness environment is the need to be perceived as an expert or the "source of truth." This is often tied to professional standards or peer-reviewed accuracy. We summarize the primary drivers across each organization type in Table 2 below.

Organization Type	Primary Driver
Important Not to Fail	Avoidance of Failure
New Ideas Valued	Pursuit of Growth
Output Correctness	Pursuit of Precision

Table 2: Organizational Type and Primary Driver

4. THE BIG FIVE PERSONALITY TRAITS

The Big Five personality traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism play a significant role in influencing how individuals interact with generative AI technologies. These traits can affect user engagement, preferences, and satisfaction with AI systems, as well as the design and functionality of AI applications. The integration of personality traits into AI systems can enhance personalization and improve user experience, as demonstrated in various studies.

- Openness: Individuals high in openness are more likely to engage with new technologies, including generative AI, due to their curiosity and willingness to explore novel experiences. This trait can lead to increased engagement and satisfaction with AI systems that offer creative and innovative solutions (Kovbasiuk et al., 2025; Arıbaş & Dağlarlı, 2024).
- Conscientiousness: This trait is associated with a preference for structured and reliable systems. AI applications that provide clear, consistent, and dependable interactions are likely to appeal to conscientious users, enhancing their engagement and satisfaction

(Priyanka et al., 2024; Cabrera-Paniagua & Rubilar-Torrealba, 2022).

- **Extraversion:** Extraverted individuals may prefer AI systems that facilitate social interaction or provide dynamic and engaging experiences. Personalized travel recommendation systems, for example, have shown higher performance rates for extraverted users (Arıbaş & Dağlarlı, 2024).
- **Agreeableness:** Users with high agreeableness may favor AI systems that are perceived as friendly and supportive. This trait can influence positive attitudes towards AI, as agreeable individuals are more likely to appreciate AI's assistance and collaboration (Babiker et al., 2024).
- **Neuroticism:** Individuals with high neuroticism may have mixed reactions to AI, potentially experiencing anxiety or skepticism. However, AI systems that offer reassurance and support can mitigate these concerns and improve user satisfaction (Babiker et al., 2024).

This paper proceeds by theorizing (Weick, 1995) about the interaction between the Big Five Personality Traits of the person using or potentially using Generative AI (Gen AI), and the three externalities, features of the organization or situation, that will be the users of the output of the person's use of Gen AI.

We do not discuss the impact of the Big 5 Personalities when the organizational focus is on the correctness of the output from Gen AI. If it is clear that the output of Gen AI is correct, then it makes sense for a person to use Gen AI whatever their Big 5 personality, since the user is able to know whether or not the output is correct, or not.

Openness to Experience (Big 5 Personality #1)

Includes: imagination, curiosity, aesthetic sensitivity, preference for novelty, and intellectual engagement.

"It's important not to fail"

- **Interaction:** People high in openness may not prioritize this construct highly — they are more comfortable with ambiguity, experimentation, and taking intellectual risks.
- **Use of Gen AI:** They might still use Gen AI for accuracy or completeness (e.g., fact-checking), but they're less motivated by the need to avoid failure, and more by exploration.

- **Risk tolerance:** High. They might be willing to "fail" as part of the creative process.

"New ideas are valued"

- **Interaction:** This is strongly aligned with Openness. These individuals are likely to actively seek out Gen AI as a tool for inspiration, brainstorming, divergent thinking, and pushing boundaries.
- **Use of Gen AI:** High engagement in tasks where novelty, creativity, or idea generation is needed. They may use Gen AI to explore possibilities, generate unique outputs, or synthesize unexpected connections.

Conscientiousness (Big 5 Personality #2)

Includes: self-discipline, organization, goal orientation, reliability, and a preference for planning over spontaneity.

"It's important not to fail"

- **Interaction:** Highly motivating. People high in conscientiousness place strong value on precision, correctness, and goal completion. Avoiding failure is central.
- **Use of Gen AI:** These users may turn to Gen AI for proofreading, planning, summarizing, or fact-checking — tasks that improve reliability and reduce the chance of error.
- **Risk tolerance:** Low. They'll use Gen AI cautiously, likely double-checking outputs.

"New ideas are valued"

- **Interaction:** Secondary. While not opposed to new ideas, people high in conscientiousness want structured, applicable innovations— ideas that can be implemented effectively.
- **Use of Gen AI:** They might use AI for process improvement or efficiency ideas but less likely for wild brainstorming.

Extraversion (Big 5 Personality #3)

Includes: sociability, assertiveness, high energy, and a tendency toward excitement, and reward-seeking behavior.

"It's important not to fail"

- **Interaction:** Less dominant. Extroverts are often comfortable with risk if there's potential for reward or recognition.
- **Use of Gen AI:** Might use Gen AI for communicative tasks like crafting persuasive emails, social media posts, or preparing for presentations — where public performance matters.

"New ideas are valued"

- Interaction: Appealing, especially if it leads to new social opportunities, excitement, or status.
- Use of Gen AI: Likely to use AI for idea generation in outward-facing tasks (e.g., event planning, team brainstorming). Might value Gen AI as a conversational partner or co-creator.

Agreeableness (Big 5 Personality #4)

Includes: compassion, cooperation, trust, and a desire to maintain social harmony.

"It's important not to fail"

- Interaction: Relevant if failure impacts others or causes conflict. Less about personal perfectionism, more about being helpful or not letting others down.
- Use of Gen AI: May use it to support or smooth social tasks — like improving clarity in writing, resolving misunderstandings, or being tactful in responses.

"New ideas are valued"

- Interaction: Mildly appealing, especially if the new ideas benefit relationships or group wellbeing.
- Use of Gen AI: Might use AI to craft thoughtful messages, generate ideas for teamwork or cooperation, or understand diverse perspectives.

Neuroticism, a.k.a. Emotional Stability, inversely (Big 5 Personality #5)

Includes: emotional sensitivity, tendency to experience anxiety, moodiness, or vulnerability to stress.

"It's important not to fail"

- Interaction: Extremely motivating. High-neuroticism individuals may fear failure intensely, often due to fear of judgment or consequences.
- Use of Gen AI: Heavy use for reassurance, checking work, validating decisions. May lean on AI as a way to reduce anxiety or feel more confident.

"New ideas are valued"

- Interaction: Could provoke anxiety unless clearly safe or structured. Might avoid novelty use unless the use is framed as low-risk or personally beneficial.
- Use of Gen AI: Lower likelihood of using Gen AI for open-ended creativity. Might prefer tasks where the AI helps them feel in control.

Personality and the organizational task focus interact in a way that impacts the use of generative AI by a person in a particular organization or situation. For instance, the "Openness" trait is tolerant of failure, and strongly aligned with ideation and brainstorming. In an organization where it is "important not to fail", being tolerant of failure means a person with a strong personality trait of openness might possibly use Gen AI even though it might hallucinate or have bias in training or output that leads to an incorrect output for the situation or organization.

We summarize the Big 5 Personalities and their interactions with the organization's task motivations leading to likely Gen AI use in Table 3 below.

Big 5 Personalities	Organizational Environment		Likely Generative AI Use
	"Important Not to Fail"	"New Ideas Are Valued"	
Openness	Tolerant of failure	Strongly aligned	Brainstorming, ideation, synthesis
Conscientiousness	Strongly avoids failure	Moderately interested if applicable	Accuracy, planning, structured innovation
Extraversion	Risk-tolerant if reward is likely	Drawn to exciting or social ideas	Communication, persuasion, collaboration
Agreeableness	Avoids failure if it harms others	Interested if it improves harmony	Diplomacy, perspective-taking, social writing
Neuroticism	Strongly avoids failure (high anxiety)	Cautious or avoids novelty	Confidence-building, checking, reassurance

Table 3: Summary of the Big 5 Personalities and how they interact with the Organizational Environments of "Important Not to Fail" and "New Ideas are Valued" leading to likely Gen AI use

Big 5 Personalities	Big 5 Level	Organizational Environment		
		Important Not to Fail	Output Correctness	New Ideas Are Valued
Openness	High	1		2
	Low	3		4
Conscientiousness	High	5		6
	Low	7		8
Extraversion	High	9		10
	Low	11		12
Agreeableness	High	13		14
	Low	15		16
Neuroticism	High	17		18
	Low	19		20

Table 4: The High/Low of the Big 5 Personalities mapped to the three contexts of “Important Not to Fail”, “Output Correctness”, and “New Ideas Are Valued”

In order to present the interaction of personality and task motivation, we use a heat map methodology. Heat maps are useful when highlighting areas of significance in data. By using color to represent varying levels of importance, heat maps allow for the quick visual identification of areas of significance within the data. When producing a heat map, how the presenter assigns a heat color is subjective. We have used just three colors: red, yellow, and green. Red means that Gen AI is not used (or should not be), yellow means Gen AI might be used, and green means that Gen AI is (or should be) used.

In the table above (Table 4), we break each of the five personality traits into high and low. High means that a person’s score on that trait after completing a personality trait assessment is in the top 30%, and low means a person’s score after completing a personality trait assessment is in the bottom 30%. This approach is consistent with the use of extreme group designs in behavioral research, which enhance interpretability and theoretical contrast when examining interaction effects (Cohen, 1983; Preacher et al., 2005). Such percentile-based categorizations are commonly used in personality research to distinguish meaningful differences in trait expression while maintaining alignment with underlying continuous constructs (John & Srivastava, 1999). Breaking the five personality traits into high and low is represented in the 10

rows of Table 1 (openness has a high and a low row, and so on for each personality trait). The table has three columns to represent each of the organizational environment task motivations. A combination of 10 rows and three columns for organizational environments offers 30 possible outcomes.

However, the column for “Output Correctness” is colored entirely green. We believe that the use of Gen AI by someone when the output can easily be seen to be correct or incorrect is not going to be moderated, or impacted, by a Big 5 Personality trait. For instance, even someone who is highly neurotic and strongly avoids failure can and should use Gen AI if they can easily see if the output is correct or incorrect. This entire column, for each of the 10 rows, is labelled green, meaning that a person should use Gen AI in this situation.

The remaining two columns, “Important not to fail” and “New ideas are valued” are covered by the following 20 heat map colors. The number here corresponds to the particular cell in the table:

1. Yellow. High openness encourages experimentation and engagement with novel technologies (McCrae & Costa, 1997; Devaraj et al., 2008), but risk-averse, failure-avoidant environments

- suppress such behavior due to the prioritization of reliability over exploration (March, 1991; Sitkin & Pablo, 1992). As a result, generative AI use remains possible but constrained, leading to conditional adoption.
2. Green. High openness to experience is associated with curiosity, experimentation, and a willingness to engage with novel technologies, which aligns strongly with innovation-oriented organizational contexts that encourage new ideas, resulting in active use of generative AI systems (McCrae & Costa, 1997; Devaraj et al., 2008; Agarwal & Prasad, 1998).
 3. Red. Low openness to experience (closed to experience) reflects a preference for routine and resistance to change, leading individuals to avoid novel technologies (McCrae & Costa, 1997; McElroy et al., 2007). In failure-avoidant organizational contexts that emphasize risk minimization and discourage experimentation (Sitkin & Pablo, 1992), this alignment reinforces avoidance of generative AI use.
 4. Yellow. Low openness to experience reflects resistance to change and low personal innovativeness in information technology (McCrae & Costa, 1997; Agarwal & Prasad, 1998; McElroy et al., 2007). In innovation-oriented organizational contexts, however, social influence and normative expectations can partially override this resistance (Venkatesh et al., 2003), leading to limited or conditional adoption of generative AI.
 5. Red. High conscientiousness is associated with reliability, structure, and adherence to rules and standards (Barrick & Mount, 1991), which can reduce engagement with uncertain or probabilistic technologies (Devaraj et al., 2008). In failure-avoidant organizational contexts that emphasize risk minimization and error avoidance (Sitkin & Pablo, 1992), this alignment leads individuals to avoid generative AI due to concerns regarding accuracy, accountability, and compliance.
 6. Yellow. High conscientiousness is associated with goal orientation, reliability, and a focus on task performance (Barrick & Mount, 1991), which promotes the use of generative AI when it is perceived as useful for improving outcomes (Davis, 1989; Devaraj et al., 2008). However, the preference for structure and control tempers this adoption in innovation-oriented environments, leading to selective and constrained use when AI introduces uncertainty.
 7. Yellow. Low conscientiousness (flexible and spontaneous) is associated with lower adherence to rules, reduced structure, and greater tolerance for ambiguity (Barrick & Mount, 1991; McElroy et al., 2007), which may increase willingness to experiment with generative AI. However, in failure-avoidant organizational contexts that emphasize risk minimization and control (Sitkin & Pablo, 1992), this creates a misalignment that leads to inconsistent or situational use rather than consistent adoption.
 8. Green. Low conscientiousness is associated with reduced adherence to rules and greater tolerance for ambiguity (Barrick & Mount, 1991; McElroy et al., 2007), which can increase willingness to experiment with new technologies and reflect higher behavioral flexibility (Agarwal & Prasad, 1998). In innovation-oriented organizational contexts that encourage new ideas, this alignment promotes active and exploratory use of generative AI.
 9. Yellow. High extraversion, characterized by sociability and assertiveness, is positively associated with engagement in interactive technologies (Amichai-Hamburger & Vinitzky, 2010; Devaraj et al., 2008). Extraverted individuals are also more responsive to social influence (Venkatesh et al., 2003), which may encourage generative AI use; however, in failure-avoidant organizational contexts that emphasize risk minimization (Sitkin & Pablo, 1992), this behavior is constrained, resulting in conditional or situational adoption.
 10. Green. High extraversion, characterized by sociability and assertiveness, is positively associated with the use of interactive and communication-oriented technologies (Amichai-Hamburger & Vinitzky, 2010; Devaraj et al., 2008). In

- innovation-oriented organizational contexts that encourage collaboration and new ideas, this alignment increases the likelihood that extraverted individuals will adopt generative AI as an interactive and augmentative tool (Venkatesh et al., 2003; Dennis et al., 2008).
11. Green. Low extraversion (introversion) is associated with a preference for solitary work and selective engagement with technology (McElroy et al., 2007). Because introverted individuals are less responsive to social influence (Venkatesh et al., 2003), generative AI adoption in failure-avoidant organizational contexts is limited; however, when the technology supports individual task efficiency, it may still be used selectively for private productivity gains (Goodhue & Thompson, 1995).
 12. Yellow. Low extraversion (introversion) is associated with reduced social engagement and lower responsiveness to social influence (McElroy et al., 2007; Venkatesh et al., 2003). Although innovation-oriented organizational contexts encourage experimentation, this reduced responsiveness dampens adoption; however, when generative AI supports individual task efficiency, it may still be used selectively, resulting in conditional use (Goodhue & Thompson, 1995).
 13. Red. High agreeableness (accommodating personality) is associated with cooperation, trust, and strong adherence to social norms (Graziano & Eisenberg, 1997), making technology use highly contingent on perceived organizational approval and peer behavior (Venkatesh et al., 2003). In failure-avoidant organizational contexts that emphasize risk minimization and discourage deviation (Sitkin & Pablo, 1992), this alignment leads to avoidance of generative AI use.
 14. Green. High agreeableness is associated with cooperation and adherence to social norms (Graziano & Eisenberg, 1997), making technology use highly contingent on perceived organizational approval and peer behavior (Venkatesh et al., 2003). In innovation-oriented organizational contexts where norms favor experimentation and new ideas (Schein, 2010), this alignment promotes the adoption of generative AI.
 15. Yellow. Low agreeableness (antagonistic personality) is associated with competitiveness, skepticism, and a reduced tendency to conform to social norms (Graziano & Eisenberg, 1997; Devaraj et al., 2008). In failure-avoidant organizational contexts that discourage risk-taking and deviation (Sitkin & Pablo, 1992), this disposition may lead to selective or covert use of generative AI, as individuals are more willing to challenge or bypass established norms when pursuing perceived performance advantages.
 16. Yellow. Low agreeableness is associated with competitiveness, independence, and a reduced tendency to conform to social norms (Graziano & Eisenberg, 1997), which can increase willingness to adopt technologies in pursuit of individual performance advantages (Devaraj et al., 2008). In innovation-oriented organizational contexts that encourage experimentation and new ideas (Schein, 2010), this alignment promotes the active and opportunistic use of generative AI.
 17. Red. High neuroticism is characterized by anxiety and heightened sensitivity to risk, which increases perceptions of uncertainty and potential negative outcomes in technology use (Thatcher & Perrewé, 2002; McElroy et al., 2007). In failure-avoidant organizational contexts that emphasize error avoidance and risk minimization (Sitkin & Pablo, 1992), this alignment leads to avoidance of generative AI adoption.
 18. Yellow. High neuroticism is characterized by anxiety and heightened sensitivity to risk, which increases perceptions of uncertainty and potential negative outcomes in technology use (Thatcher & Perrewé, 2002; McElroy et al., 2007). In innovation-oriented organizational contexts that encourage experimentation, social influence and supportive norms can partially offset these concerns (Venkatesh et al., 2003), leading to hesitant or conditional use of generative AI rather than complete avoidance.

19. Red. Low neuroticism (emotionally stable personality) is associated with confidence, low anxiety, and tolerance for ambiguity, which increases willingness to engage with uncertain technologies such as generative AI (Thatcher & Perrewé, 2002; Devaraj et al., 2008). This confidence is consistent with higher perceived self-efficacy in technology use (Compeau & Higgins, 1995). However, in failure-avoidant organizational contexts that emphasize risk minimization (Sitkin & Pablo, 1992), this willingness is constrained, resulting in conditional or situational adoption.
20. Green. Low neuroticism (emotional stability) is associated with confidence, low anxiety, and a high tolerance for ambiguity, which increases willingness to engage with uncertain technologies such as generative AI (Thatcher & Perrewé, 2002; Devaraj et al., 2008). In innovation-oriented organizational contexts that encourage experimentation and new ideas, this alignment facilitates active and sustained use of generative AI (Venkatesh et al., 2003; March, 1991).

An important thing to consider here is that this is a framework development paper. There has been no research that has examined the interaction between personality types and organizational tasks in Gen AI environment. The heatmap here is essentially a list of hypotheses, with our color-coding representing what we think will be the outcome of actual testing.

5. PUTTING IT ALL INTO PRACTICE

The framework we have presented here operates at the individual level. Each person has a particular set of scores for the Big Five Personalities, and it is a person's individual personality that interacts with the with the organizational environment and its task motivations, and determines their personal type of generative AI use.

Putting this framework to work to understand human-AI interaction is a three-step process.

Step 1 is for the individual to take the Big Five Personality test, and to record the scores from that. A useful Big Five Personality test is available here:

<https://www.123test.com/personality-test/>

One of the authors of this paper took the test and the scores are as follows:

1. Openness to experience: 46 = This is too close to mid-range to be a strong personality, so we ignore this trait
2. Conscientiousness: 71 = "high"
3. Extraversion: 93 = "high"
4. Agreeableness: 2 = "low"
5. Neuroticism: 6 = "low"

The five personality traits of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism each now align with a particular row in the assessment chart.

Step 2 of the process is to use these scores and their high/mid/low level range and apply the level to each organizational environment per Table 4. We summarize the author's results in Table 5.

Big 5 Personalities	Big 5 Level	Organizational Environment		
		Important Not to Fail	Output Correctness	New Ideas Are Valued
Openness	Mid-range	Not scored	Not scored	Not scored
Conscientiousness	High			
Extraversion	High			
Agreeableness	Low			
Neuroticism	Low			

Table 5: results of one of the author's Big 5 Personality test

Step 3 is to interpret the results for each of the organizational environment task motivation columns. If you are at a particular type of organization, as described above, then you may only need to see the results for that particular column.

As seen in Table 5, if the organization or situation is "*Important not to fail*": two reds and two yellows means that this person should not, and probably does not, use Gen AI here. Red for "conscientiousness" is because this personality trait strongly avoids failure. High extraversion is yellow as a person with this personality trait is risk-tolerant if reward is likely. That is, Gen AI use may be tolerated by the person even in a situation where it is important not to fail provided that the reward of success is high enough. The low Agreeableness (antagonistic personality trait) is assertive, skeptical, critical, independent, and non-conformist. Whereas the "Important Not to Fail" environment (externality) is rules-focused, risk-averse, error-intolerant, outcome-precise. As such, this is a person-environment mismatch, and the low agreeable (antagonistic) individual may experience tension or constraint in how they use Gen AI. Low neuroticism (emotionally stable) means that you use Gen AI if it's interesting, but it is more "cool curiosity" than passionate exploration, and with the organizational environment focused on avoidance of failure, Gen AI use is shown as red from Table 4. In sum, in an organization where it is important not to fail, this author would likely not use Gen AI except if the rewards for using it were high enough.

As seen in Table 5, if the organization or situation is such that "*New ideas are valued*": three yellows and one green suggest that this person could easily be using Gen AI in this situation, but it is not a given. High "conscientiousness" means that accuracy and planning are important, so Gen AI use is limited to structured tasks that focus on this. High extraversion means that Gen AI use is likely for idea generation in outward-facing tasks (e.g., event planning, team brainstorming), and it's also likely that Gen AI might be valued as a co-creator. Low agreeableness, (antagonistic personality) can be an asset in idea-valuing environments, especially when independent thinking is rewarded and constructive dissent is welcome, but can become a liability when collaboration or group cohesion are essential. Agreeableness is a yellow as this person probably uses Gen AI to test assumptions or debunk norms, uses Gen AI as a solo ideation partner, not as a consensus tool, but is less likely to use Gen AI for team brainstorming, shared documents, or communication polishing. Low neuroticism

(emotionally stable) in this context means that the person is unafraid to experiment with new or risky prompts, and can bounce back easily from failure or poor outputs. The innovation-oriented environment encourages ideation, and hence this is shown as green for Gen AI use.

6. DISCUSSION

As expressed earlier, this is a theory development paper. We have provided instructions on how to use the theory in the preceding "Putting it all together" section. In addition, in Appendix A we provide three further examples of colleagues who have put the theory to the test. We acknowledge that this sample size is too small to determine the validity of our theory. As a theory development paper, we aim at establishing a conceptual framework for understanding the intersection of personality, organizational environment, and generative AI usage. While the heat maps and preliminary data provided here offer valuable initial insights into these relationships, the current sample size is intended for exploratory "proof of concept" purposes rather than definitive empirical validation. Consistent with the iterative nature of Information Systems research, this study focuses on the formulation of the theoretical model. Subsequent research will involve a large-scale, longitudinal study designed to statistically validate the framework and its generalizability across more diverse organizational contexts.

7. CONCLUSION

The idea of developing a framework that shows the interaction of the Big 5 Personalities with different types of organizational environment task motivation is useful in at least three meaningful ways:

1. This framework helps predict how people actually use Gen AI in real-world contexts. Most studies on AI use focus on features or outcomes, but this paper focuses on the psychology of the user. This matters because, a) people don't use AI the same way even when doing the same task, and b) the framework connects personality, task motivation, and tool behavior, which explains why users act differently. This finding could lead to better personalized AI training, guidance, and adoption strategies in schools, workplaces, and creative industries. A potential use case here is that a company training employees on Gen AI could tailor onboarding differently for low-Openness staff (who want clarity and structure) than for

high-Openness staff (who want creative play).

2. This framework offers a model for predicting and shaping Gen AI adoption. The framework can be used to predict who will use AI (and for what tasks), explain why some people avoid it, and suggest how to design interfaces, prompts, or education differently based on user traits. A potential use case is that a university could better support Gen AI in writing centers if they understood that low-Conscientiousness students might not use AI unless it's embedded into the workflow.
3. This framework opens the door to interventions that make AI use more equitable and effective. Right now, AI use is often haphazard. Some students overuse it, some avoid it, some misuse it. This framework helps design behavioral nudges ("Try this for brainstorming!"), custom prompts based on personality, and support tools for those less likely to engage creatively or confidently. For instance, an adaptive learning tool could adjust its prompts based on the user's Big Five profile — helping low-Neuroticism users take more creative risks, or giving low-Openness users clearer scaffolding.

In short, this paper is not just saying "People use Gen AI differently." Instead, we propose a framework to answer the question of who uses Gen AI, how, and why they are using it based on who they are and what the organization or situation is looking to achieve. Thus, our findings have academic, practical, and design value. What is needed next is to test the framework and the "hypotheses" that the heatmap contains.

8. REFERENCES

- Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, 9(2), 204–215. <https://doi.org/10.1287/isre.9.2.204>
- Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289–1295. <https://doi.org/10.1016/j.chb.2010.03.018>
- Arıbaş, E., & Dağlarlı, E. (2024). Transforming personalized travel recommendations: Integrating generative AI with personality models. *Electronics*, 13(23), 4751. <https://doi.org/10.3390/electronics13234751>
- Babiker, A., Alshakhsi, S., Supti, T. I., & Ali, R. (2024). Do personality traits impact the attitudes towards artificial intelligence? In *Proceedings of the 11th International Conference on Behavioral and Social Computing (BESC)* (pp. 1–8). IEEE. <https://doi.org/10.1109/BESC64747.2024.10780777>
- Barczyk, C., & Duncan, D. (2017). Facebook enhanced college courses and the impact of personality on sense of classroom community. *Information Systems Education Journal*, 15(1), 42–54. <http://isedj.org/2017-15/>
- Barnett, T., Pearson, A. W., Pearson, R., & Kellermanns, F. W. (2015). Five-factor model personality traits as predictors of perceived and actual usage of technology. *European Journal of Information Systems*, 24(4), 374–390. <https://doi.org/10.1057/ejis.2014.10>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Cabrera-Paniagua, D., & Rubilar-Torrealba, R. (2022). Adaptive intelligent autonomous system using artificial somatic markers and Big Five personality traits. *Knowledge-Based Systems*, 249. <https://doi.org/10.1016/j.knosys.2022.108995>
- Carr, N. G. (2003). IT doesn't matter. *Harvard Business Review*, May 2003, 41–49.
- Chau, K. F. (2024). Understanding mature students' intention to use mobile learning in higher education: Integrating personality traits and UTAUT. *International Journal of Educational Management*, 38(4), 1062–1078. <https://doi.org/10.1108/IJEM-01-2023-0008>
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249–253. <https://doi.org/10.1177/014662168300700301>

- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189–211. <https://doi.org/10.2307/249688>
- Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, 6(4), 343–359. <https://doi.org/10.1521/pedi.1992.6.4.343>
- 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>
- Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, 32(3), 575–600. <https://doi.org/10.2307/25148857>
- Devaraj, S., Easley, R. F., & Crant, J. M. (2008). Research note—How does personality matter? Relating the five-factor model to technology acceptance and use. *Information Systems Research*, 19(1), 93–105. <https://doi.org/10.1287/isre.1070.0153>
- Drucker, P. F. (1994). The theory of the business. *Harvard Business Review*, September–October, 95–104.
- Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19(2), 213–236. <https://doi.org/10.2307/249689>
- Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 30(3), 611–642. <https://doi.org/10.2307/25148742>
- Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 795–824). Academic Press.
- Hignite, M. A., Satzinger, J. W., & Margavio, T. M. (1999). Correlated factors of success: Personality, creativity and academic achievement among IS students. *Journal of Information Systems Education*, 10(3–4), 12–20.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). Guilford Press.
- John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & Oliver P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York: Guilford Press
- Joshi, A. R., Das, S., & Sekar, S. (2023). How Big Five personality traits affect ICT use: A meta-analysis. *Australasian Journal of Information Systems*, 27. <https://doi.org/10.3127/ajis.v27i0.3985>
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87(3), 530–541. <https://doi.org/10.1037/0021-9010.87.3.530>
- Kovbasiuk, A., Triantoro, T., Przegalińska, A., Sowa, K., Ciechanowski, L., & Gloor, P. (2025). The personality profile of early generative AI adopters: A Big Five perspective. *Central European Management Journal*, 33(2), 252–264. <https://doi.org/10.1108/cemj-02-2024-0067>
- Labti, A., & Bourhim, E. M. (2022). How does personality matter? Investigating the effect of Big Five personality traits on VR acceptance behavior. In *ICCCNT Proceedings*. <https://doi.org/10.1109/ICCCNT54827.2022.9984605>
- Liu, J., & Chen, J. (2025). Chatbot-aided product purchases among Generation Z: The role of personality traits. *Frontiers in Psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1454197>
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87. <https://doi.org/10.1287/orsc.2.1.71>
- McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516. <https://doi.org/10.1037/0003-066X.52.5.509>

- McCrae, R. R., & Costa Jr, P. T. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research*, 2(1999), 139-153.
- McElroy, J. C., Hendrickson, A. R., Townsend, A. M., & DeMarie, S. M. (2007). Dispositional factors in internet use: Personality versus cognitive style. *MIS Quarterly*, 31(4), 809-820. <https://doi.org/10.2307/25148821>
- Prayoga, T., & Abraham, J. (2016). Behavioral intention to use Instagram: The role of personal innovativeness and Big Five personality traits.
- Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, 10(2), 178-192. <https://doi.org/10.1037/1082-989X.10.2.178>
- Priyanka, S., Santhiya, S., Keerthika, S., Divyadharshini, B., Madan, P., & Yaswin, M. (2024, June). Personality Based Text Generation. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICCCNT61001.2024.10724424>
- Reynolds, J. H., Adams, D. R., Ferguson, R. C., & Leidig, P. M. (2017). The personality of a computing major. *Information Systems Education Journal*, 15(4), 20-28.
- Schein, E. H. (2010). *Organizational culture and leadership* (4th ed.). Jossey-Bass.
- Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review*, 17(1), 9-38. <https://doi.org/10.5465/amr.1992.4279564>
- Svendsen, G. B., Johnsen, J. A. K., Almås-Sørensen, L., & Vittersø, J. (2013). Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology*, 32(4), 323-334. <https://doi.org/10.1080/0144929X.2011.553740>
- Thatcher, J. B., & Perrewé, P. L. (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly*, 26(4), 381-396. <https://doi.org/10.2307/4132314>
- Tripathi, N., Upadhyay, S., & Das, H. (2022). Forced transition to technology: Role of self-efficacy and big five personality variables in the adoption of technology. *International Journal of Educational Management*, 36(7), 1131-1151. <https://doi.org/10.1108/IJEM-02-2022-0065>
- 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>
- Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, 40(3), 385-390. <https://doi.org/10.2307/2393789>
- Woszczyński, A. B., Guthrie, T. C., & Shade, S. (2005). Personality and programming. *Journal of Information Systems Education*, 16(3), 293-299.

APPENDIX A

Limited Testing of the Framework

Three colleagues took the Big 5 Personality test referenced in our study. Below, we provide a quantitative list of their results, a heat map of their results across organizational environments, and a discussion of how their responses fit with the framework described in our study. These examples underscore the practical use of our *Personality-Organizational Environment* framework.

Colleague 1

With a Big 5 Personality result of:

1. Openness to experience: 39 = This is too close to mid-range to be a strong personality
2. Conscientiousness: 83 = "High"
3. Extraversion: 83 = "High"
4. Agreeableness: 81 = "High"
5. Neuroticism: 44 = This is too close to mid-range to be a strong personality so ignore

Big 5 Personalities	Big 5 Level	Organizational Environment		
		Important Not to Fail	Output Correctness	New Ideas Are Valued
Openness	Mid-range	39 out of 100	39 out of 100	39 out of 100
Conscientiousness	High			
Extraversion	High			
Agreeableness	High			
Neuroticism	Mid-range	44 out of 100		44 out of 100

Table 1: results of Colleague 1’s Big 5 Personality test summarized

For Colleague 1, we note that if their organizational environment is "Important not to fail" then their results indicate two reds and one yellow in the heat map. This colleague likely won't use Gen AI much at all in an important not to fail environment. If their organizational environment is "new ideas are valued" then their heat map would reflect one yellow and two greens. In a "New ideas are valued" scenario, Colleague 1 is likely leaning into using Gen AI as much as possible, as the organizational environment is supportive of new ideas and is innovation-oriented. High conscientiousness is associated with goal orientation, reliability, and a focus on task performance (Barrick & Mount, 1991), which promotes the use of generative AI when it is perceived as useful for improving outcomes (Davis, 1989; Devaraj et al., 2008). However, the preference for structure and control tempers this adoption in innovation-oriented environments, leading to selective and constrained use when AI introduces uncertainty.

Colleague 2

With a Big 5 Personality result of:

1. Openness to experience: 12 = "low"
2. Conscientiousness: 17 = "low"
3. Extraversion: 38 = This is too close to mid-range to be a strong personality so ignore
4. Agreeableness: 32 = This is too close to mid-range to be a strong personality so ignore
5. Neuroticism: 42 = This is too close to mid-range to be a strong personality so ignore

Big 5 Personalities	Big 5 Level	Organizational Environment		
		Important Not to Fail	Output Correctness	New Ideas Are Valued
Openness	Low			
Conscientiousness	Low			
Extraversion	Mid-range	38 out of 100		38 out of 100
Agreeableness	Mid-range	32 out of 100		32 out of 100
Neuroticism	Mid-range	42 out of 100		42 out of 100

Table 2: results of Colleague 2’s Big 5 Personality test summarized

For Colleague 2, we note that if their organizational environment is "Important not to fail" then their results indicate two yellows. In this environment, Colleague 2 likely won't use Gen AI much at all. If their organizational environment is "new ideas are valued" then Colleague 2’s results are two greens. High openness to experience is associated with curiosity, experimentation, and a willingness to engage with novel technologies, which aligns strongly with innovation-oriented organizational contexts that encourage new ideas, resulting in active use of generative AI systems (McCrae & Costa, 1997; Devaraj et al., 2008; Agarwal & Prasad, 1998). Low conscientiousness is associated with reduced adherence to rules and greater tolerance for ambiguity (Barrick & Mount, 1991; McElroy et al., 2007), which can increase willingness to experiment with new technologies and reflect higher behavioral flexibility (Agarwal & Prasad, 1998). In innovation-oriented organizational contexts that encourage new ideas, this alignment promotes active and exploratory use of generative AI.

Colleague 3

With a Big 5 Personality result of:

1. Openness to experience: 87 = "high"
2. Conscientiousness: 81 = "high"
3. Extraversion: 95 = "high"
4. Agreeableness: 46 = This is too close to mid-range to be a strong personality so ignore
5. Neuroticism: 14 = "low"

Big 5 Personalities	Big 5 Level	Organizational Environment		
		Important Not to Fail	Output Correctness	New Ideas Are Valued
Openness	High			
Conscientiousness	High			
Extraversion	High			
Agreeableness	Mid-range	46 out of 100		46 out of 100
Neuroticism	Low			

Appendix Table 3: results of Colleague 3’s Big 5 Personality test summarized

For Colleague 3, we note that if their organizational environment is "Important not to fail" then their results indicate two reds and two yellows. This colleague appears tolerant of failure (yellow for "high openness to experience"), strongly avoids failure (red for "high conscientiousness"), risk-tolerant if reward is likely (yellow for "high extraversion"), and strongly avoids failure (red for "low neuroticism", emotionally stable). With two reds and two yellows, the "reward is likely" would have to be very strong to overcome the two personality traits of "strongly avoids failure".

If Colleague 3's organizational environment is "new ideas are valued", then their high "openness to experience" means that they will likely use Gen AI for brainstorming and ideation. Colleague 3's high "conscientiousness" means that accuracy and planning are important, so their Gen AI use is limited to structured tasks that focus on this. With a high "extraversion", Colleague 3 will likely use Gen AI for communication, persuasion, and collaboration. This colleague's low "neuroticism" (emotionally stable personality) means that they likely use Gen AI if it's interesting, but with the organization being innovation-oriented and rewarding such, this pushes Colleague 3 to green for Gen AI use.

APPENDIX B

Description of the Big 5 Personality Traits

The following chart offers a detailed description of high- and low-level Big 5 personality traits:

Big 5 Personality	Description of the Personality Trait	
	High Level	Low Level
Openness (to Experience)	Characterized by curiosity, creativity, and a preference for novelty and variety. Individuals high in openness are more likely to seek out new experiences, embrace innovation, and engage with abstract or complex ideas (Costa & McCrae, 1992; John et al., 2008).	Reflects a preference for routine, familiarity, and practical approaches. Individuals low in openness tend to be more conservative, resistant to change, and less inclined to experiment with new ideas or approaches (Costa & McCrae, 1992; McElroy et al., 2007).
Conscientiousness	Associated with organization, reliability, discipline, and goal-oriented behavior. Individuals high in conscientiousness are structured, detail-oriented, and motivated to achieve high standards of performance (Barrick & Mount, 1991; John et al., 2008).	Characterized by lower levels of organization and discipline. Individuals low in conscientiousness may be more flexible and spontaneous but can also be less reliable, less structured, and more prone to inconsistency in task completion (Barrick & Mount, 1991; McElroy et al., 2007).
Extraversion	Defined by sociability, assertiveness, and a preference for interaction with others. Individuals high in extraversion are energized by social engagement and are often enthusiastic, talkative, and outwardly expressive (Costa & McCrae, 1992; Amichai-Hamburger & Vinitzky, 2010).	Reflects introversion, with a preference for solitary activities and lower levels of social interaction. Individuals low in extraversion tend to be more reserved, reflective, and focused on independent work (John et al., 2008; McElroy et al., 2007).
Agreeableness	Characterized by cooperation, trust, empathy, and a tendency to prioritize social harmony. Individuals high in agreeableness are supportive, collaborative, and inclined to follow social norms (Graziano & Eisenberg, 1997; John et al., 2008).	Associated with competitiveness, skepticism, and a lower tendency to conform to social expectations. Individuals low in agreeableness may be more independent, critical, and willing to challenge others or established norms (Graziano & Eisenberg, 1997; Devaraj et al., 2008).
Neuroticism (Emotional Stability)	High neuroticism reflects emotional sensitivity, anxiety, and a tendency to experience negative emotions such as stress or worry. Individuals high in neuroticism may be more risk-averse and reactive to uncertainty (Costa & McCrae, 1992; Thatcher & Perrewé, 2002).	Low neuroticism (high emotional stability) is characterized by calmness, resilience, and confidence. Individuals low in neuroticism are more likely to remain composed under pressure and tolerate ambiguity or uncertainty (John et al., 2008; Devaraj et al., 2008).

Adapting CAMIL for K–12 World Language Learning: A Qualitative Study of Immersive VR in Middle School Spanish

Emily Loughlin
Emily.Loughlin@gmail.com

Kevin Mentzer
Kevin.Mentzer@Nichols.edu

Nichols College
Dudley, MA 01571

Abstract

This exploratory study examines how immersive virtual reality (IVR) can support world language acquisition in middle school classrooms. Seventh- and eighth-grade Spanish students used head-mounted display (HMD) headsets with the ImmerseMe platform to practice conversational interactions in simulated, authentic settings. To interpret learners' experiences, the study applied the Cognitive Affective Model of Immersive Learning (CAMIL), which positions presence and agency as key affordances shaping affective and cognitive outcomes. CAMIL was adapted for adolescent learners by organizing six factors (interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation) into three developmentally appropriate domains. Data were collected through student surveys and teacher observations and analyzed thematically using the adapted model. Students reported high levels of social and physical presence, moderate motivation, and increased confidence in pronunciation. However, limited agency within the platform and intermittent technical barriers reduced engagement and, at times, contributed to cognitive overload. Overall, findings indicate that IVR can strengthen engagement and perceived learning in middle school language instruction when presence is paired with meaningful learner agency and when implementation reduces friction and supports autonomy. This study contributes an adapted CAMIL framework for younger learners and offers design and deployment implications for IVR in K–12 education.

Keywords: Immersive virtual reality (IVR); world language learning; middle school; CAMIL; presence.

Recommended Citation: Loughlin, E., Mentzer, K., (2026). Adapting CAMIL for K–12 World Language Learning: A Qualitative Study of Immersive VR in Middle School Spanish. *Information Systems Education Journal*, v25(n1) pp 55-65. DOI# <https://doi.org/10.62273/DIWX3099>

Adapting CAMIL for K–12 World Language Learning: A Qualitative Study of Immersive VR in Middle School Spanish

Emily Loughlin and Kevin Mentzer

1. INTRODUCTION

Virtual reality (VR) is emerging as a promising tool in education because it can simulate real-world experiences in ways that deepen attention, participation, and experiential understanding. Across disciplines, VR has been used to support learning outcomes ranging from technical skill development to exploratory engagement with complex environments, and its adoption in K–12 settings continues to grow.

Recent research has also extended VR's use into higher education business and communication courses, particularly through head-mounted display (HMD) platforms that support experiential learning. For example, a multi-site ISCAP study used Meta Quest headsets to simulate business scenarios and found that immersion increased student engagement, although implementation challenges (e.g., setup time and limited content control) constrained outcomes (Mentzer et al., 2025). Similarly, VR has been applied in public speaking instruction to strengthen learner confidence and promote self-awareness in delivery, suggesting broader potential for communication-based learning contexts (Patterson et al., 2025).

As VR hardware and software mature, educators are increasingly exploring immersive virtual reality (IVR), especially HMD-based IVR, as a viable instructional approach across content areas. World language education is a particularly strong fit for IVR because language learning depends on context, interaction, and multisensory cues. Traditional classroom instruction often struggles to provide authentic linguistic and cultural exposure. In contrast, IVR can place learners in simulated real-world settings where they can practice listening and speaking, engage in contextualized conversations, and explore cultural environments in safe, controlled spaces. This affordance may be especially valuable in middle school classrooms, where students are building foundational skills and typically have limited access to native-speaking environments or study abroad experiences.

Despite growing interest in IVR for learning,

relatively little is known about how HMD-based IVR functions in middle school settings, particularly for novice language learners. Much of the existing literature focuses on adult learners or technical training contexts, leaving an important gap in understanding how adolescent learners experience immersion in language learning. In addition, commonly used immersive learning frameworks have rarely been adapted to reflect the cognitive, emotional, and behavioral characteristics of younger learners.

To address this gap, the present study examines HMD-based IVR in seventh- and eighth-grade Spanish classrooms using the Cognitive Affective Model of Immersive Learning (CAMIL). CAMIL, originally developed for adult learners, identifies presence and agency as key psychological affordances that shape affective and cognitive learning outcomes. This study adapts CAMIL to better reflect adolescent learning characteristics and applies the revised framework to qualitative data gathered from student surveys and classroom observations.

Beyond its pedagogical focus, this study also aligns with increasing expectations for business schools to demonstrate societal impact through community partnerships and educational outreach. AACSB's business accreditation standards emphasize mission-aligned engagement that improves stakeholder outcomes and supports broader societal goals (AACSB, 2023). By partnering with a middle school to explore immersive language instruction, this project illustrates one pathway for business faculty to contribute to educational impact while advancing research on emerging instructional technologies.

The sections that follow review the relevant literature and describe the adapted theoretical framework, then detail the classroom implementation and methodology. Results are organized according to the revised CAMIL constructs, and the paper concludes with implications for instructional design, platform development, and future research on immersive learning in secondary education.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

Immersive virtual reality (IVR) has gained increasing attention as an instructional technology because it can place learners in simulated environments that support experiential, contextualized learning. Across educational settings, IVR has been associated with increased engagement, focused attention, and opportunities for active participation. At the same time, researchers caution that these benefits depend heavily on effective instructional design and may be constrained by usability issues, cognitive overload, cost, accessibility, and health concerns such as simulator sickness (Cabrera-Duffaut et al., 2024; Kaimara et al., 2022).

World language learning is a particularly relevant context for IVR because language acquisition depends on meaningful interaction, situational context, and cultural cues. Traditional classroom instruction often provides limited opportunities for authentic communicative practice, especially for novice learners. IVR can help address this limitation by situating students in simulated real-world settings where they can practice listening and speaking in context, engage with culturally situated prompts, and experiment with language in a low-risk environment. Prior research suggests that virtual environments can support intercultural communication, vocabulary development, and learner engagement in second-language instruction (DeWitt et al., 2022; Legault et al., 2019; Lin & Lan, 2015).

Although the literature points to the promise of IVR for language learning, much of this work has focused on adult or higher education populations. Less is known about how head-mounted display (HMD) IVR functions in middle school settings, where learners may differ developmentally in attention, self-regulation, and comfort with immersive technologies. In addition, immersive learning frameworks have rarely been adapted specifically for adolescent learners in K-12 environments. This gap is especially important because younger learners may experience the motivational and cognitive demands of immersive environments differently than adults.

This study is guided by the Cognitive Affective Model (see Figure 1) of Immersive Learning (CAMIL) developed by Makransky and Petersen (2021). CAMIL explains how immersive technologies influence learning through the interaction of technological affordances and learner-centered psychological processes. The

model identifies two central affordances, presence and agency, that shape six related affective and cognitive factors: interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation. According to CAMIL, immersive learning environments are most likely to support learning when these affordances and learner conditions are aligned in ways that promote meaningful engagement rather than distraction or overload.

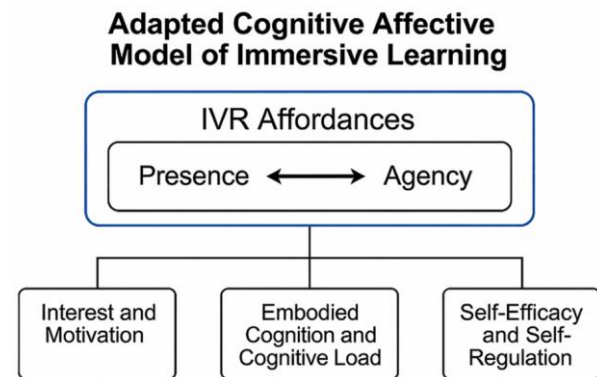


Figure 3: Adapted CAMIL Model

Presence refers to the subjective sense of "being there" in a virtual environment. Prior scholarship conceptualizes presence including physical presence, social presence, and self-presence (Ijsselsteijn & Riva, 2003; Lee, 2004; Lombard & Ditton, 1997). Physical presence reflects the feeling of being located within the virtual space. Social presence refers to the sense of interacting with others in that space, including avatars or AI-driven agents. Self-presence refers to the perception of oneself as an active participant in the environment. Together, these dimensions help explain why immersive environments can feel more immediate and personally involving than traditional digital interfaces.

Agency refers to the learner's perceived ability to act within and influence the virtual environment. When learners feel that they can control pacing, make meaningful choices, and affect the flow of an experience, they are more likely to report engagement, autonomy, and confidence (Moore & Fletcher, 2012). In immersive learning contexts, agency is especially important because presence alone does not guarantee productive learning. A virtual environment may feel vivid and absorbing, but if learners perceive little control over navigation, difficulty, or interaction, the benefits of immersion may be reduced.

In addition to presence and agency, CAMIL highlights six affective and cognitive factors that mediate learning in immersive environments.

Interest and motivation shape whether learners are drawn into the task and remain engaged over time. Embodiment reflects the extent to which cognition is experienced as situated within action and sensory experience. Cognitive load refers to the mental effort required to process information and navigate the environment, which can support or hinder learning depending on the quality of the instructional design. Self-efficacy and self-regulation relate to learners' confidence, persistence, and ability to manage attention and effort during the learning process. Together, these factors provide a useful framework for examining how immersive experiences influence both engagement and perceived learning.

Prior studies suggest that IVR can support these dimensions when learners are able to interact with realistic environments, receive timely feedback, and engage in focused, contextualized practice. For example, research has shown that immersive environments can increase engagement through multisensory and game-like experiences while also supporting vocabulary learning and intercultural communication (DeWitt et al., 2022; Legault et al., 2019). At the same time, the literature also shows that immersive experiences may become less effective when learners encounter technical friction, excessive sensory stimulation, or limited control over pacing and task difficulty (Çoban et al., 2024; Freina & Ott, 2015; Kaimara et al., 2022; Javadi et al., 2024). These concerns may be especially relevant for younger learners, who may require more scaffolding and clearer structures for self-regulation.

Health and usability concerns are also important in K-12 contexts. Studies have noted that children and adolescents may be particularly sensitive to issues such as dizziness, visual discomfort, disorientation, and cognitive fatigue when using immersive systems (Kaimara et al., 2022). In educational settings, these concerns are compounded by practical implementation challenges such as headset calibration, battery management, onboarding time, and the need for close monitoring during use. These factors suggest that the effectiveness of IVR depends not only on the quality of the virtual content but also on the usability and classroom fit of the technology itself.

Because CAMIL was developed primarily in adult learning contexts, this study adapts the model for middle school language learners. The adapted version retains the central importance of presence and agency while consolidating the six original learner factors into three

developmentally appropriate domains: interest and motivation; embodied cognition and cognitive load; and self-efficacy and self-regulation. Figure 1 presents the adapted CAMIL model used in this study. This adaptation offers a more practical framework for examining adolescent learners' experiences while preserving CAMIL's core theoretical logic. Grouping the constructs in this way is intended to reflect the interrelated nature of affective and cognitive responses in middle school classrooms, where engagement, confidence, and self-management often operate together rather than as fully distinct processes.

This adaptation is also appropriate for world language learning because language acquisition is both cognitive and social. Learners must attend to vocabulary, pronunciation, context, and cultural meaning while also managing uncertainty and taking interpersonal risks. In immersive environments, these demands may be intensified by the sensory richness and novelty of the medium. As a result, a framework that accounts for both the affective and cognitive dimensions of learning is especially useful for understanding how students experience IVR in language instruction.

Taken together, the literature suggests that IVR has strong potential for world language learning because it can provide contextualized, interactive practice that is difficult to replicate in traditional classrooms. However, the effectiveness of IVR appears to depend on the balance between immersive affordances and instructional usability. Presence may increase engagement and authenticity, but meaningful learning is more likely when learners also experience agency, manageable cognitive demands, and sufficient support for confidence and self-regulation. To examine these dynamics in a middle school Spanish context, the present study applies an adapted CAMIL framework to explore how presence, agency, and related affective and cognitive factors shape engagement and perceived learning outcomes.

3. METHODOLOGY

This section describes the study design, participants, instructional context, and procedures for data collection and analysis. The implementation of immersive virtual reality (IVR) in a middle school Spanish classroom was examined using an exploratory qualitative approach guided by the adapted Cognitive Affective Model of Immersive Learning (CAMIL). Data were collected from student reflections and

classroom observations to examine learners' affective and cognitive responses to head-mounted display (HMD) VR.

This study was guided by the following research question: "How does immersive virtual reality (IVR), through the affordances of presence and agency as conceptualized in an adapted CAMIL framework, influence engagement and perceived learning outcomes among middle school Spanish learners?"

Research Design

This study used an exploratory qualitative design to investigate how IVR supports language learning among middle school students. Guided by CAMIL, the analysis focused on how presence and agency shaped engagement, motivation, and perceived learning outcomes. CAMIL was adapted for this context to better reflect adolescent learners by consolidating the original six cognitive and affective factors into three domains: interest and motivation; embodied cognition and cognitive load; and self-efficacy and self-regulation.

Participants and Setting

The study was conducted in spring 2025 at a public K–8 school in rural central Massachusetts. Participants included 30 seventh-grade students and 38 eighth-grade students enrolled in Spanish classes taught across multiple sections by the same teacher. Students were novice learners in their first or second year of Spanish. Participation in the IVR activities occurred during regular instruction and was not tied to additional assessment or grading. Students used school-issued Chromebooks in class and could choose to use the VR platform.



Figure 4: Classroom Setting

Technology and Platform

Students used Meta Quest 3 headsets (<https://www.meta.com/quest/quest-3/>) to access ImmerseMe (<https://immerseme.co/>), a commercial VR platform designed to support

guided language conversations in immersive, culturally situated environments. The platform allowed students to select scenarios such as dining in a restaurant or exploring a marketplace and to practice scripted conversations with recorded characters (Aros et al., 2024; Briand, 2021; Karageorgakis, 2022). While ImmerseMe offered some flexibility in pacing and scenario selection, student interaction was largely limited to responding to scripted prompts.

Headsets were rotated among students. Twenty headsets were loaned to the school, each assigned a city name. Because the teacher taught five class sections, each device was typically shared by five students. Additional headsets were available when technical issues occurred. Devices were stored securely and sanitized regularly; however, battery drainage and controller alignment occasionally interfered with use. Students received instruction on device handling and platform navigation, and many were using HMD-based VR for the first time.

Instructional Design

The IVR activities were integrated into a broader unit on conversational Spanish. Over ten weeks, students completed approximately 10 to 20 IVR sessions, typically lasting 30 minutes or less. The teacher selected ImmerseMe lessons aligned with classroom objectives (for example, greetings, food and drink, and body parts), and students were encouraged to self-select difficulty levels or scenarios within those constraints.

Across lessons, students could choose among speaking, typing, translating, or listening activities based on comfort level. Although the platform was intended to provide opportunities for low-stress speaking practice, student responses varied due to individual differences in comfort, cognitive load, and prior exposure to immersive technologies.

Data Collection

Data were collected through two rounds of open-ended student questionnaires administered via Google Forms, once in May and once in June. The May questionnaire captured initial perceptions, while the June questionnaire served as a follow-up to assess changes over time. The teacher noted that many students declined to complete the follow-up due to end-of-year fatigue and reported that they would respond only if their opinions had changed. As a result, the May responses served as the primary dataset for analysis.

Additional qualitative data were collected through

researcher field notes and classroom observations, including informal conversations with students about their experiences. No formal recordings were used. The researcher documented observations related to headset use, off-task behavior, engagement, and technical issues. These notes provided context to supplement student reflections. The teacher also maintained observation notes during and after headset use and added end-of-day reflections on lesson implementation.

Data Analysis

Survey responses and observation notes were coded thematically using the adapted CAMIL framework. Responses were grouped into five constructs: presence; agency; interest and motivation; embodied cognition and cognitive load; and self-efficacy and self-regulation. Representative student quotes were selected to reflect both common themes and divergent perspectives and were anonymized by grade level and gender.

Ethical Considerations

The study received Institutional Review Board (IRB) approval through Nichols College. Parental consent and student assent were obtained in accordance with ethical research guidelines. Responses were anonymous, and no audio, video, or screen recordings were collected. Students could opt out of VR use or switch to the Chromebook version of the platform if they experienced discomfort.

4. RESULTS

Findings from student surveys and teacher observations are presented in alignment with the study's research question, examining how presence and agency, as framed within the adapted CAMIL model, influence engagement and perceived learning outcomes. Student feedback was analyzed thematically across five key constructs: presence, agency, interest and motivation, embodied cognition and cognitive load, and self-efficacy and self-regulation. Direct quotations from students are included to illustrate common perceptions, challenges, and variation in experience.

Overview of Data Collection

Students completed an open-ended qualitative questionnaire twice, once in May and again in June. The June dataset was less robust due to survey fatigue, end-of-year disengagement, and the instruction that students should complete the second survey only if their opinions had changed. Approximately 90% of students opted not to

complete the June survey. Most responses remained stable. One notable exception involved a student who initially enjoyed the VR experience but later described it as "repetitive and annoying."

A recurring concern involved speech recognition. Several students reported that classroom background noise affected microphone performance. In many cases, the headset microphones captured student speech, but ImmerseMe did not consistently convert speech to accurate text. Students with speech impediments were especially frustrated by this issue, and many students reported at least one instance where the platform did not recognize what they said.

Presence

Student responses indicated varied experiences with presence. Most students described social presence through simulated interactions within the ImmerseMe platform. One male eighth grader noted the value of "speaking to a real person," which increased his confidence. Students also described physical presence in comments such as:

- "While I was using it, it was like I was actually there" – Female, Grade 8
- "When we were in the cities it made it feel so real" – Male, Grade 7

Self-presence was evident when students referenced their avatars or described their appearance in VR. While many reported feeling immersed, others suggested that limited interaction with the environment reduced their overall sense of presence.



Figure 3: Students Actively Engaged

Overall, these findings align with CAMIL's emphasis on presence as a key affordance of immersive virtual reality. Students described both physical and social presence, which likely contributed to engagement and perceived immersion.

Agency

Students described limited agency within the ImmerseMe platform. Several commented on the lack of control over the environment and the limited ability to tailor difficulty. One eighth grader noted:

- "I can never find a lesson that is in my perfect level in Spanish. It is either too hard or too easy."

Another student suggested that a more game-like structure could increase engagement and motivation:

- "If the VR was set up so it's like a quest game... it could motivate the learning experience." – Male, Grade 8

Although the platform offered some choice within lessons, students could not navigate freely or dynamically adjust settings, which constrained perceived agency. While many students reported feeling immersed and "actually there" within the VR environments, their experience of agency often constrained the potential benefits of presence. One student shared,

- "I like to use VR to learn Spanish because I can go at my own pace, but... it's a lot of time I'm stuck online and I would rather not spend my whole day behind screens,"

reflecting how limited control over content flow and platform design may temper the motivational benefits of high presence. Another remarked,

- "I like using them, the only downside is how long it takes to get in"

These comments underscore that presence alone is insufficient. Without meaningful agency, students may become disengaged or frustrated, even in richly immersive environments.

Consistent with CAMIL, limited agency constrained students' ability to navigate and control their experience, moderating the positive effects of presence. This reinforces the importance of balancing both affordances to sustain engagement and support learner autonomy.

Interest and Motivation (Combined)

Interest and motivation were strongly expressed by most students, especially those who had not previously used VR. For many, the novelty and personal relevance of the lessons contributed to engagement:

- "I found the topics that were more personal or aimed towards me the most helpful, because they were more interesting." – Female, Grade 8

Students appreciated working at their own pace and being in a different environment:

- "I like using VR to learn Spanish because I can go at my own pace." – Female, Grade 8
- "It makes class even more interesting... you can challenge yourself or just practice at the level you feel comfortable on." – Male, Grade 8

However, some voiced concerns about the tedium of repetition, technical barriers to entry, or overstimulation. One student noted:

- "It was very quick and I didn't get a lot of time to process the words." – Female, Grade 7

Despite frustrations, many learners expressed a desire to continue using the tool, provided the usability improves.

This combination of interest and motivation aligns with CAMIL's assertion that affective engagement increases when students experience high presence and a sense of personal relevance. However, diminished agency and technical challenges appeared to temper this engagement over time.

Embodied Cognition and Cognitive Load (Combined)

Students reported feeling fully immersed in the language-learning context, often drawing comparisons to real-life conversations or travel experiences:

- "It feels like I'm having a full Spanish conversation in the real world where I will actually use Spanish." – Male, Grade 8
- "It helps to capture my attention, since the only thing I can see is the lesson." – Male, Grade 8

That said, cognitive load emerged as a challenge for some. A handful of students described issues with simulator sickness, blurry visuals, or overstimulation:

- "I struggled with staying in the headset because it made me feel woozy." – Female, Grade 8
- "I don't like VR because it gives me motion sickness." – Male, Grade 8

These cognitive limitations, while not universal, suggest the importance of limiting headset time and offering alternative platforms when needed.

As outlined in CAMIL, immersive sensory input can enhance embodiment and deepen conceptual engagement. At the same time, cognitive overload, particularly from discomfort or system limitations, can reduce learners' ability to fully benefit from these experiences.

Self-Efficacy and Self-Regulation (Briefly Mentioned)

While these were less directly addressed in survey responses, some students reported feeling more confident after using ImmerseMe:

- "I feel like since I've started [VR lessons], my pronunciation of Spanish is getting much better." – Male, Grade 8

The teacher observed that most students followed through on their tasks and stayed engaged. However, because the researcher lacked visibility into what students were doing inside the headsets, verifying on-task behavior was difficult. Notably, while some students joked about using the devices for games, no off-task behavior was confirmed.

Although these factors were less overtly referenced, student persistence and self-monitoring reflect the role of self-regulation in immersive environments. This is consistent with CAMIL, which links strong regulation to lower cognitive load and improved learner confidence.

Taken together, these findings highlight the interdependence of presence and agency as conceptualized in the adapted CAMIL model. Their balance appears important for sustaining motivation and supporting cognitive regulation in immersive learning environments.

5. CONCLUSIONS AND FUTURE WORK

This study examined how immersive virtual reality (IVR), through the affordances of presence and agency as conceptualized in an adapted Cognitive Affective Model of Immersive Learning (CAMIL), can support world language instruction in middle school Spanish classrooms. The adapted model consolidated six CAMIL factors

into three domains appropriate for K–12 learners: interest and motivation; embodied cognition and cognitive load; and self-efficacy and self-regulation.

Overall, most students responded positively to the immersive experience. Students reported strong social and physical presence, high engagement, and enjoyment, and many valued the ability to work at their own pace. Several students also reported increased confidence in pronunciation. At the same time, student feedback indicated that agency within the ImmerseMe platform was limited, and technical frustrations and overstimulation sometimes interfered with deeper learning and sustained motivation. Cognitive barriers, including simulator sickness, poor visual clarity, and constrained control over the learning path, reduced learning efficacy for some students.

These findings support the relevance of CAMIL in secondary education. When presence and agency are balanced, IVR can support language practice, cultural exposure, and learner autonomy. Outcomes appear to depend less on novelty and more on thoughtful instructional design, pacing, and usability.

This work also illustrates how research with middle school learners can advance both scholarly and institutional priorities. By leveraging immersive learning expertise to enhance world language instruction in K–12 classrooms, the project provides a model of community engagement that supports meaningful stakeholder outcomes.

Future Work

This exploratory study suggests several directions for future research. First, future studies should use tools that allow instructors to view student activity within the headset in real time to strengthen observation validity. Second, adding quantitative measures such as pre and post assessments of vocabulary retention or oral proficiency would help validate and extend the perceptual findings reported here. Comparative studies across traditional instruction, video-based activities, and immersive instruction could also clarify which outcomes are uniquely supported by IVR. Finally, platform features such as gamified elements, differentiated learning paths, and improved speech recognition may strengthen agency and sustain motivation over time.

Limitations

Several limitations should be considered. The study was conducted in a single school with one

teacher, which may limit generalizability and introduce instructional bias. Findings relied primarily on qualitative surveys and observations rather than formal proficiency measures. Technical constraints, limited platform functionality, limited teacher visibility into headset activity, and occasional discomfort may also have shaped engagement. In addition, follow-up data collection was reduced by end-of-year fatigue, limiting longitudinal comparison. Despite these limitations, the study contributes evidence supporting the adapted CAMIL framework and identifies practical considerations for IVR integration in K–12 language learning.

6. ACKNOWLEDGMENTS

Generative AI was used to assist with language refinement and readability improvement during preparation of the final manuscript draft. The authors reviewed and revised all content and take full responsibility for the originality, accuracy, and integrity of the submission.

The contents of this research paper were developed under grant #P116Z230123 from the U.S. Department of Education.

7. REFERENCES

- AACSB. (2023). AACSB and Societal Impact: Aligning with the AACSB 2020 Business Accreditation Standards. Tampa, FL: AACSB International.
- Aros, M., Tyger, C.L., Chaparro, B.S. (2024). Unraveling the Meta Quest 3: An Out-of-Box Experience of the Future of Mixed Reality Headsets. In: Stephanidis, C., Antona, M., Ntoa, S., Salvendy, G. (eds) HCI International 2024 Posters. HCII 2024. Communications in Computer and Information Science, vol 2116. Springer, Cham. https://doi.org/10.1007/978-3-031-61950-2_1
- Briand, C. (2021, May 21). Embedding ImmerseMe into classroom practice and curriculum design. ImmerseMe - Using virtual reality to learn languages. <https://immerseme.co/blog/2021/5/21/chloe-briand>
- Cabrera-Duffaut, A., Pinto-Llorente, A. M., & Iglesias-Rodríguez, A. (2024). Immersive learning platforms: Analyzing virtual reality contribution to competence development in higher education—a systematic literature review. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1391560>
- Çoban, M., Topuz, A. C., Karabulut, A., & K. F. Chiu, T. (2024). Comparing realities: A study on the impact of virtual reality versus paper-based reading on higher education students' comprehension skills. *Participatory Educational Research*, 11(4), 19-36. <https://doi.org/10.17275/per.24.47.11.4>
- DeWitt, D., Chan, S.F. & Loban, R. (2022). Virtual reality for developing intercultural communication competence in Mandarin as a Foreign language. *Education Tech Research Dev* 70, 615-638. <https://doi.org/10.1007/s11423-021-10074-9>
- Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *The International Scientific Conference eLearning and Software for Education 1*(33), 10-1007. <https://doi.org/10.12753/2066-026x-15-020>
- Ijsselsteijn, W., & Riva, G. (2003). Being there: The experience of presence in mediated environments. In G. Riva, F. Davide, & W. . Ijsselsteijn (Eds.), *Being There: Concepts, effects and measurement of user presence in synthetic environments* (pp. 1–14). Ios Press.
- Javadi, E., Gebauer, J., & Jung, K. (2024). *Investigating the potential of virtual reality medium for collaborative work. Results from a series of pilot studies*. Proceedings of the ISCAP Conference. <https://iscap.us/proceedings/2024/pdf/6136.pdf>
- Kaimara, P., Oikonomou, A. & Deliyannis, I. (2022) Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns. *Virtual Reality* 26, 697–735. <https://doi.org/10.1007/s10055-021-00563-w>
- Karageorgakis, T. (2019, October 16). ImmerseMe: Using virtual reality to learn languages — Educraft. Educraft. <https://educraft.tech/immerseme/>
- Lee, K. M. (2004). Presence, explicated. *Communication theory*, 14(1), 27-50. <https://doi.org/10.1111/j.1468-2885.2004.tb00302.x>
- Legault, J., Zhao, J., Chi, Y. -A., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages*, 4(1), 13.

- <https://doi.org/10.3390/languages4010013>
- Lin, T.J. & Lan, K. (2015). Language Learning in Virtual Reality Environments: Past, Present, and Future. *Educational Technology & Society*. 18. 486-497.
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2), JCMC321. <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>
- Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937-958. <https://doi.org/10.1007/s10648-020-09586-2>
- Mentzer, K., Frydenberg, M., & Li, S. (2025). Exploring VR-enhanced learning in business education: A multi-site study. *Information Systems Education Journal*, 23(5), 30-43. <https://doi.org/10.62273/WOFR1838>
- Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: a review of cue integration approaches. *Consciousness and cognition*, 21(1), 59-68. 10.1016/j.concog.2011.08.010
- Patterson, A., Temple, C., Anderson, N., Rogalski, C., & Mentzer, K. (2025). The Virtual Stage: Virtual Reality integration in effective speaking courses. *Information Systems Education Journal*, 23(4), 57-68. <https://doi.org/10.62273/IMLU5983>

APPENDIX A

VR TPS Nichols Qualitative post-research questions

VR Specific Questions

1. Have you used VR headsets before this experience in Spanish class?
 - a. If so, how would you describe your experience, and for what did you use the VR headset?
2. How do you feel about using technology like VR in your learning process overall?
3. Were there any parts of the VR experience that you found challenging or difficult? Can you explain?
4. How do you think the use of VR could be adjusted to better support your learning?
5. What kind of activities or elements in the VR experience did you find most engaging or helpful?
6. How do you think the use of VR might be useful for learning other subjects or skills?
7. In what ways do you think VR can be improved to better support your learning experience?
8. Do you believe that using VR in language learning will be helpful for other subjects or areas of study? Why?

Spanish Specific Questions

1. What was your initial reaction when you first used a Virtual Reality headset in your Spanish class?
 - a. What did you like about it?
 - b. What did you dislike?
2. How did using the VR headset help you engage with the Spanish language compared to traditional learning methods in class?
3. In what ways did the use of VR influence your involvement or interest in the Spanish language?
4. Can you describe a specific experience or activity in the VR environment that helped you understand Spanish?
5. How do you feel that VR technology influenced your motivation to practice or learn Spanish outside of class?
6. In your opinion, how effective is VR in helping you develop your Spanish skills?
7. Were there any challenges or difficulties you encountered while using the VR headset to learn Spanish?
 - a. If so, how did you overcome them?
8. Do you think VR technology can help you feel more confident in speaking Spanish? Why or why not?
9. How do you think using VR to practice Spanish will impact your ability to understand and communicate in real-life situations?
10. What would you tell a friend about using VR to learn Spanish? Would you recommend it? Why or why not?
11. How do you think your perception of learning languages has changed after using VR technology in your Spanish class?

Developing an Artificial Intelligence Course for a Small Undergraduate Program

Scott Turner
scott.turner@richmond.edu

Lionel Mew
lmew@richmond.edu

Terry Turner
terry.turner@richmond.edu

University of Richmond
Richmond, VA 23173 USA

Abstract

This paper presents the systematic design and development of an undergraduate artificial intelligence and machine learning course intended to serve both technical and non-technical students in higher education. The work addresses growing industry demand for AI and ML competencies by proposing a comprehensive course framework that accommodates students from diverse academic backgrounds while maintaining academic rigor. Building upon established experiential and active learning theories, the proposed course design emphasizes hands-on learning through progressive skill development. The curriculum incorporates fundamental concepts including supervised and unsupervised learning, neural networks, natural language processing, and computer vision, while integrating ethical considerations throughout. The pedagogical framework utilizes cloud-based laboratory environments and industry-standard tools to provide accessible yet rigorous learning experiences that bridge theoretical understanding with practical implementation skills. This pedagogical design case synthesizes current best practices in AI and ML education, drawing from successful data analytics program implementations to develop a comprehensive framework for course design.

Keywords: Artificial Intelligence, Machine Learning, Course Development, Curriculum Design, Undergraduate Education, Experiential Learning

Recommended Citation: Turner, S., Mew, L., Turner, T.G., (2026). Developing an Artificial Intelligence Course for a Small Undergraduate Program. *Information Systems Education Journal*, v25(n1) pp 66-79. DOI# <https://doi.org/10.62273/HWXF5405>

Developing an Artificial Intelligence Course for a Small Undergraduate Program

Scott Turner, Lionel Mew, and Terry Turner

1. INTRODUCTION

The Growing Demand for Artificial Intelligence and Machine Learning Education

The rapid proliferation of artificial intelligence (AI) and machine learning (ML) technologies across industries has created unprecedented demand for skilled professionals who can design, implement, and deploy intelligent systems. According to the World Economic Forum's Future of Jobs Report 2025, artificial intelligence is projected to drive the creation of 170 million new roles globally by 2030, while displacing 92 million — a net gain of 78 million jobs (World Economic Forum, 2025). This explosive growth has pressured higher education institutions to develop comprehensive educational programs preparing students for careers in these rapidly evolving fields.

The challenge facing academic institutions extends beyond meeting market demand. AI and ML education requires unique blends of theoretical understanding, practical implementation skills, and ethical reasoning capabilities. Unlike traditional computer science or mathematics courses, AI and ML programs must prepare students to work with complex, evolving technologies while developing critical thinking skills necessary for navigating intelligent systems' societal implications. This multifaceted educational challenge has prompted institutions worldwide to reconsider traditional pedagogical approaches and develop innovative curriculum designs effectively preparing students for the AI-driven future.

Educational Challenges in AI and ML Instruction

Contemporary AI and ML education faces distinct challenges differentiating it from other technical disciplines. First, rapid technological advancement requires curricula balancing foundational concepts with cutting-edge developments exposure. Technologies like large language models, generative AI, and advanced neural architectures continue evolving at unprecedented pace, requiring educational programs to remain current while maintaining academic rigor.

Second, AI and ML education must accommodate students from diverse academic backgrounds. While traditional computer science students possess strong programming and mathematical foundations, AI applications' interdisciplinary nature attracts students from business, psychology, biology, and social sciences. This diversity necessitates pedagogical approaches providing sufficient technical depth while remaining accessible to learners with varying technical preparation levels.

Third, AI and ML competencies' practical nature requires extensive hands-on learning experiences. Students must understand theoretical concepts like gradient descent, neural network architectures, and statistical learning theory, while developing proficiency by applying these concepts using real-world datasets and industry-standard tools. This practical experience requirement creates logistical challenges related to computing resources, software licensing, and project design.

The Need for Systematic Course Development

Despite growing AI and ML educational programs, limited research exists on systematic course development approaches specifically tailored to these fields. While broader technical education literature provides valuable insights, AI and ML's unique characteristics—including interdisciplinary nature, rapid technological evolution, and societal implications—require specialized pedagogical consideration and careful design decisions.

Most existing AI and ML education literature focuses on program descriptions or isolated teaching techniques rather than comprehensive course development frameworks (Chen et al., 2020; Coursera, 2024; Ouyang & Jiao, 2021; Zawacki-Richter et al., 2019). This gap represents significant opportunity for institutions to benefit from systematic approaches integrating pedagogical theory with practical implementation considerations. Developing well-designed AI and ML course frameworks can serve as blueprints for institutions seeking to establish or enhance offerings in these critical areas.

Institutional Context and Program Integration

Our institution's School of Professional and Continuing Studies (SPCS) serves diverse traditional and non-traditional students, with average age 37 and 81% enrolled part-time. This demographic diversity provides ideal context for developing AI and ML courses serving both technically oriented students seeking advanced skills and working professionals understanding AI applications in their fields.

The proposed course was designed as part of expanding data analytics offerings, building upon our successful undergraduate data analytics program implementation (Mew, 2020; Clayton & Clopton, 2019). This foundation provides valuable insights into effective technical education pedagogical approaches in our institutional context, including strategies for accommodating diverse student preparations, balancing theoretical depth with practical accessibility, and creating meaningful learning experiences within evening and weekend class schedule constraints.

This study presents systematic design and development of an undergraduate AI and ML course framework addressing these challenges through a pedagogical design approach, synthesizing insights from technical education literature, successful related field implementations, and established learning theories. Our approach employs backwards design principles and emphasizes practical application while maintaining theoretical foundations necessary for professional competency. As a pedagogical design case, this work presents a framework grounded in a single institutional context; empirical validation of learning outcomes represents an important direction for future research.

2. LITERATURE REVIEW

Theoretical Foundations for AI and ML Education

Systematic development of effective AI and ML courses requires careful consideration of pedagogical theory, curriculum design principles, and practical implementation strategies. This literature review synthesizes existing research to inform undergraduate AI and ML course design, focusing on theoretical foundations, curriculum structure considerations, and assessment approaches.

Kolb's (1984) experiential learning theory provides fundamental framework for designing effective AI and ML courses, given the hands-on

nature required for developing competency in these fields. The theory's four-stage learning cycle—concrete experience, reflective observation, abstract conceptualization, and active experimentation—aligns naturally with machine learning's iterative development process, where students engage with data, observe patterns, conceptualize models, and experiment with implementations.

Technical education applications demonstrate particular relevance for AI and ML instruction. McCarthy (2010) emphasizes that "learning by doing" is essential for developing technical competencies and critical thinking skills necessary for addressing complex intelligent systems problems. This theoretical foundation suggests effective AI and ML courses should prioritize hands-on experiences over purely theoretical instruction, directly informing curriculum design decisions.

In our course design, we specifically applied Kolb's cycle by structuring laboratory experiences where students first engage with real datasets (concrete experience), analyze output patterns (reflective observation), learn underlying algorithms (abstract conceptualization), and then modify code for their own industry applications (active experimentation). This cyclical approach directly shaped our four-lab sequence and portfolio assessment structure.

Active learning theory (Felder & Brent, 2009; Prince, 2004) provides crucial guidance for designing AI and ML courses that move beyond passive information consumption to engage students in analytical thinking and problem-solving. The theory's emphasis on two-way communication, higher-order thinking, and experiential involvement directly addresses the complex skill set required for AI and ML competency.

Bonwell and Eison (1991) identify key active learning characteristics particularly relevant to AI and ML instruction: student involvement in activities beyond listening, emphasis on developing skills rather than transmitting information, and engagement in higher-order thinking processes like analysis, synthesis, and evaluation. These characteristics align closely with competencies required for effective AI and ML practice, suggesting active learning approaches should be central to course design.

Curriculum Design and Content Integration

A critical AI and ML course design challenge involves determining appropriate prerequisite

knowledge and foundational concept integration strategies. AI and ML's multidisciplinary nature requires students to develop mathematics, statistics, programming, and domain-specific application competencies, creating complex content sequencing and skill development decisions.

Technical education research suggests effective AI and ML courses must balance mathematical rigor with practical accessibility. Sharp Sight Labs (2016) argues that while academic AI research requires extensive mathematical preparation, entry-level industry positions emphasize effective tool and technique use rather than algorithm development from first principles. This distinction impacts undergraduate course design, suggesting courses should prioritize conceptual understanding and practical application over mathematical derivations.

Integrating mathematical concepts into AI and ML courses requires attention to pedagogical sequencing. Rather than requiring extensive mathematical prerequisites, successful programs embed necessary mathematical concepts within applied contexts. This approach allows students to develop mathematical understanding through practical application. This approach, demonstrated in successful data analytics programs (Mew, 2020; Zhang et al., 2020), suggests AI and ML courses can accommodate diverse student backgrounds while maintaining technical rigor.

Effective AI and ML curriculum design requires careful content progression building complexity gradually while maintaining engagement (Dean, 2020). Literature suggests organizing principles including beginning with supervised learning providing clear feedback, progressing to unsupervised learning requiring abstract thinking, incorporating deep learning after students master foundational algorithms, and integrating ethical considerations throughout rather than separately.

Technology Integration and Infrastructure

Tool and technology selection significantly impacts AI and ML instruction effectiveness. Literature reveals tension between providing industry-standard tool experience and ensuring accessibility for diverse technical backgrounds. Successful programs employ platforms providing powerful capabilities while minimizing programming barriers.

Zhang (2020) and Liu & Burns (2018) demonstrate effectiveness of platforms like RapidMiner enabling students without extensive

programming experience to engage advanced analytical techniques. This approach allows courses to focus on conceptual understanding and practical application rather than syntax and debugging, making AI and ML concepts accessible to broader populations while maintaining technical rigor.

Cloud-based laboratory environments offer particular advantages by providing consistent computing resources, eliminating software installation barriers, and enabling access to powerful hardware for computationally intensive tasks. The shift toward cloud-based educational infrastructure reflects practical necessities and pedagogical opportunities, allowing instructors to focus on learning objectives rather than technical logistics (Mew, 2015).

Software tool selection and licensing requires careful consideration of educational objectives, student backgrounds, and sustainability. Many institutions succeed with combinations of open-source tools providing transparency and flexibility, cloud-based platforms offering powerful capabilities with minimal setup, and industry-standard tools providing professional relevance.

Assessment and Evaluation Strategies

AI and ML competencies' complex nature requires assessment strategies beyond traditional examinations to evaluate technical skills and broader capabilities like problem-solving, ethical reasoning, and communication. Effective AI and ML assessment should include practical projects demonstrating technical competency, written assignments requiring explanation and justification of design decisions, presentations developing communication skills, and peer evaluations providing multiple perspectives on student work.

Recent AI-enhanced assessment developments offer new possibilities for evaluating student learning. AI can automatically grade certain assignments while providing immediate feedback, track student progress through learning analytics, and identify students needing additional support (Hooda, 2022). However, AI tools in assessment raise questions about academic integrity and the need for human judgment in evaluating complex work.

The practical orientation of AI and ML skills makes competency-based evaluation frameworks particularly appropriate for course assessment. Rather than focusing on knowledge recall, these approaches evaluate students' ability to apply

concepts in realistic contexts, solve complex problems, and demonstrate professional capabilities. Effective competency-based assessment typically includes portfolio development demonstrating skill progression, capstone projects integrating multiple competencies, and professional skill development including communication and teamwork.

Project-based learning emerges as particularly effective for AI and ML instruction, providing opportunities to apply concepts in realistic contexts while developing problem-solving and communication skills. Effective project design principles include addressing authentic problems that motivate engagement, incorporating iterative development cycles mirroring professional practice, requiring design decisions with justifications, and culminating in presentations developing communication skills.

3. DISCUSSION

Course Design Framework Summary: Integration of Theoretical Foundations

The course design built upon Kolb's (1984) experiential learning theory and active learning principles, emphasizing hands-on experience with AI/ML tools and iterative skill development through practical application rather than passive instruction. This integration of theoretical foundations created engaging educational experiences that mirror professional AI/ML workflows and development processes, where students engage with data, observe patterns, conceptualize models, and experiment with implementations.

The course employed backwards design methodology (Wiggins & McTighe, 2005) starting with desired workplace outcomes including stakeholder meeting participation, AI vocabulary mastery, and technology understanding, then systematically developing assessments and learning activities to achieve these professional readiness goals. Content followed a "building blocks" approach with prompt engineering established as the foundational problem-solving skill enabling subsequent learning across diverse AI/ML applications.

Technology Platform Selection and Infrastructure

Technology platform selection emphasized entirely free tools including Google Colab for Jupyter notebooks, GitHub for industry-standard repositories, and free-tier AI platforms to eliminate cost barriers while providing industry-relevant experience with tools used by software

developers and cybersecurity professionals. This approach ensured broad institutional accessibility while maintaining professional relevance and preparing students for real-world collaborative development environments.

The cloud-based infrastructure eliminated traditional barriers including software licensing, hardware specifications, and technical support complexity while providing consistent computing resources for diverse student populations. Students gained experience with these tools while developing security consciousness essential for professional practice through GitHub repository management and code sharing protocols.

Assessment Strategy Alignment

Assessment strategy emphasized hands-on portfolio development weighted at 60% using these tools, industry-specific discussion questions at 20% that prevented generic responses while encouraging field-relevant exploration, and time-limited quizzes at 20% that reinforced lesson attendance and comprehension. This multi-modal approach evaluated technical competencies alongside communication skills and ethical reasoning capabilities.

The portfolio development component required students to create professional-quality GitHub repositories demonstrating progressive skill acquisition through four structured laboratory experiences. Industry-specific discussion questions required students to connect technical concepts with their professional contexts, preventing academic dishonesty while encouraging authentic exploration of personally relevant applications.

Addressing Key Design Challenges: Balancing Technical Depth with Accessibility

The course adopted a "foot deep and half a mile wide" approach to balance technical depth with accessibility for diverse backgrounds, covering essential concepts without mathematical derivations while focusing on enterprise vocabulary and applications students would encounter in stakeholder meetings. This approach effectively served students from complete beginners and experienced professionals seeking to understand AI/ML applications in their fields.

Rather than requiring extensive mathematical prerequisites, the course embedded necessary conceptual understanding within practical contexts, allowing students to develop technical literacy through hands-on application. This strategy proved particularly effective for

accommodating non-traditional students with diverse professional backgrounds while maintaining sufficient technical rigor for career advancement.

Managing Rapid Technological Change

To manage rapid technological change in curriculum design, the course implemented modular structure with stable foundational content and adaptable emerging technology sections that could be updated regularly without disrupting overall course coherence. This design enabled continuous curriculum evolution while maintaining pedagogical soundness and learning objective alignment.

The modular approach allowed for real-time incorporation of current developments, including new AI tools, industry applications, and ethical considerations emerging from rapidly evolving technological landscape. Faculty could update specific modules based on daily news developments and industry innovations while preserving the overall course structure and progression.

Integrating Ethical Considerations

Ethical considerations were embedded within technical contexts rather than treated as separate units, requiring students to consider bias, fairness, and societal impact as integral components of AI/ML implementation decisions. This integration ensured ethical reasoning became fundamental to professional practice rather than optional academic content, preparing graduates to participate responsibly in AI development and deployment.

Students examined ethical implications throughout technical instruction, including data collection and preparation biases, algorithmic fairness in model development, transparency and explainability requirements, and societal impact considerations in deployment decisions. This approach fostered critical thinking about responsible AI development while building technical competencies.

Pedagogical Innovations and Contributions: AI-Powered Content Customization

A primary innovation involved developing systematic approaches for AI-powered content customization, enabling rapid generation of business scenarios, case studies, and examples tailored to individual student backgrounds and career interests. This methodology allowed near-individualized instruction within group settings and quick content revision that would have been

prohibitively time-intensive using traditional manual approaches.

The AI-powered customization enabled instructors to create industry-specific examples, incorporate current events and developments, and adapt content to specific cohort characteristics within hours or minutes of class delivery. This capability represented significant advancement over traditional static curriculum approaches, providing personalized learning experiences while maintaining efficient instruction delivery.

Systematic Design Option Evaluation

The course development process systematically evaluated three distinct design approaches during development: narrow focus emphasizing machine learning algorithms for business applications, broad coverage addressing comprehensive AI/ML enterprise applications, and split format dividing instruction evenly between AI and ML topics. The broad coverage approach was selected to effectively serve students while providing comprehensive enterprise exposure necessary for professional stakeholder participation.

This systematic evaluation process provided evidence-based rationale for design decisions rather than relying on intuitive preferences or traditional academic structures. The analysis demonstrated how comprehensive enterprise exposure better prepared students for professional roles requiring AI/ML literacy across diverse organizational contexts.

Progressive Skill-Building Methodology

Progressive skill-building through problem-solving mastery established prompt engineering as foundational capability enabling students to independently navigate complex topics and troubleshoot coding challenges throughout subsequent learning experiences. This approach built confidence and autonomy in technical problem-solving while providing transferable skills applicable across diverse AI/ML applications.

The prompt engineering foundation enabled students to effectively utilize AI tools for learning support, code debugging, and concept exploration, creating self-directed learning capabilities essential for keeping current with rapidly evolving technologies. Students developed metacognitive skills for approaching unfamiliar technical challenges and leveraging available resources effectively.

Industry-Agnostic Laboratory Design

Industry-agnostic laboratory design implemented four structured labs requiring students to select their own industry contexts and datasets, preventing academic dishonesty while encouraging authentic exploration of personally relevant professional applications. This approach enabled assessment of technical competencies while accommodating diverse student backgrounds and career interests.

The laboratory structure provided guided skill development through code-along methodology while requiring students to apply concepts within their chosen professional contexts. This design prevented generic responses while ensuring all students developed core competencies regardless of their specific industry applications or career trajectories.

Implementation Considerations and Practical Implications: Institutional Readiness and Resource Requirements

Technology infrastructure requirements were minimized to reliable internet access and free Google accounts for cloud-based tools, eliminating traditional barriers of software licensing, hardware specifications, and technical support complexity. This approach ensured broad institutional accessibility while providing industry-relevant experience with professional-grade tools and platforms.

Faculty development needs included formal AI/ML coursework, hands-on project experience with industry tools, and ongoing engagement with technological developments to maintain currency and practical understanding. The framework required instructor comfort with AI content generation tools and willingness to dynamically revise materials based on current developments and student needs.

Student support services established no formal prerequisites while providing prompt engineering skills early to support independent problem-solving, accommodating diverse technical backgrounds through accessible entry points and progressive skill development. Support systems addressed both academic learning needs and professional development objectives through career-relevant portfolio building.

Scalability and Adaptability

The framework demonstrated potential for adoption across diverse institutional contexts due to flexible technology requirements, modular content structure, and accommodation strategies

for varying student backgrounds and scheduling constraints. Core pedagogical principles remained applicable while allowing local adaptation to institutional resources and student characteristics.

Modifications for varied backgrounds could be achieved through industry-specific examples and discussion topics while maintaining consistent learning objectives and assessment strategies across different professional backgrounds. The framework provided sufficient flexibility for adaptation while preserving essential educational components and professional preparation objectives.

Integration with existing programs positioned the course to complement data analytics programs and serve as foundation for advanced AI/ML coursework while providing standalone professional development value (MIT, 2024). This integration strategy supported both traditional degree pathways and continuing education objectives, maximizing institutional resource utilization and student pathway options.

Broader Implications for AI/ML Education: Contributions to Educational Theory

The course demonstrated effective application of Kolb's learning cycle to machine learning workflows, where students engage with data, observe patterns, conceptualize models, and experiment with implementations. This application validated experiential learning theory's relevance for AI/ML instruction and provided concrete evidence for hands-on learning effectiveness in technical education.

Evidence for active learning effectiveness included frequent discussion questions, hands-on exploration, and collaborative problem-solving as effective engagement strategies for maintaining attention and comprehension during extended technical sessions. The approach demonstrated particular value for non-traditional students requiring flexible learning modalities and practical application focus.

Industry and Workforce Preparation

The course aligned with current industry skill demands by addressing enterprise needs for AI/ML literacy across diverse professional roles, (Berkeley, 2024) emphasizing stakeholder communication and collaborative skills alongside technical competencies. The approach prepared graduates for leadership roles in AI/ML implementation rather than purely technical development positions.

Preparation for emerging career pathways included entrepreneurial components and automation consulting opportunities while building portfolio materials demonstrating practical competencies to potential employers. Students developed both technical implementation abilities and strategic thinking skills necessary for professional advancement in AI-driven organizations.

Development of both technical and professional competencies integrated communication skills, ethical reasoning, and collaborative problem-solving with technical implementation abilities, preparing graduates for leadership roles in AI/ML implementation. This comprehensive skill development addressed industry needs for professionals capable of bridging technical capabilities with business objectives and organizational change management.

Future Research Directions

Future research should prioritize empirical evaluation of the proposed course framework through systematic assessment of student learning outcomes, skill retention, and career preparation effectiveness compared to traditional AI/ML education approaches. Controlled studies comparing different pedagogical methods would validate educational investment and provide evidence for framework effectiveness across diverse contexts.

Longitudinal studies should track graduate career advancement, professional AI/ML utilization patterns, and continued learning behaviors to validate educational approach effectiveness and career preparation impact. Research should examine whether course graduates demonstrate superior professional preparation compared to traditional program alumni and identify specific competencies contributing to career success.

Comparative implementation studies should investigate framework adaptation success across diverse institutional contexts, examining how different resource environments, student populations, and organizational cultures affect course effectiveness and implementation challenges. Research should identify critical success factors and adaptation strategies for various educational settings.

Development of standardized assessment instruments for AI/ML competencies requires creation of validated evaluation tools measuring complex capabilities including technical skills, ethical reasoning, communication competencies, and professional application abilities. These

instruments would support institutional improvement efforts and enable meaningful comparison across different educational approaches.

4. CONCLUSIONS

Synthesis of Course Development Framework

This study successfully developed a comprehensive, theory-based course framework that integrates experiential and active learning theories with AI/ML instruction while balancing technical rigor with accessibility for diverse populations. The systematic approach addressed rapid technological evolution through modular curriculum structure with stable foundational components and adaptable sections for emerging technologies, providing replicable methodology for institutions facing similar educational challenges.

Key design innovations included progressive skill-building methodology from supervised to unsupervised learning, multi-modal project structure using guided laboratory experiences, embedded ethical considerations throughout technical instruction, and cloud-based infrastructure for scalable, accessible implementation. The AI-powered content customization represented significant advancement in technical education, enabling personalized learning experiences while maintaining efficient instruction delivery.

The framework successfully addressed the critical challenge of serving different preparation levels including complete beginners and experienced professionals through flexible content delivery, personalized examples, and multiple pathway support strategies. The backwards design implementation beginning with workplace outcomes ensured professional relevance while maintaining academic rigor and theoretical foundations.

Contributions to AI/ML Education Literature

The framework contributes evidence-based approaches for applying established learning theories to emerging technical fields and provides comprehensive blueprint for institutional course adoption and implementation. The systematic evaluation of design options and detailed documentation of implementation considerations address current literature gaps where most AI/ML education reports focus on program descriptions rather than comprehensive development frameworks.

Theoretical contributions include demonstrated effectiveness of experiential learning applications in AI/ML contexts and validation of active learning approaches for maintaining engagement during extended technical sessions. The practical model for integrating ethics into technical education offers replicable methodology for other technical disciplines facing similar integration challenges.

The comprehensive development methodology synthesizing industry knowledge, academic theory, and practical implementation experience provides template for systematic course development in rapidly evolving disciplines. This approach addresses the need for evidence-based frameworks supporting institutional decision-making and resource allocation for emerging technology education.

Implications for Educational Practice and Industry Preparation

The course addresses urgent industry demand for AI/ML professionals while preparing ethically-minded practitioners for responsible AI development. By integrating communication skills, ethical reasoning, and collaborative problem-solving alongside technical competencies, the framework prepares graduates for leadership roles capable of bridging technical capabilities with business objectives — addressing workforce development needs while contributing to responsible AI advancement across diverse industry sectors.

Limitations and Future Research Priorities

Primary limitations include single-institution perspective requiring adaptation for different settings and need for empirical validation of educational effectiveness. The framework requires systematic evaluation of learning outcomes, career preparation effectiveness, and long-term professional impact to validate pedagogical approaches and institutional investment decisions.

The dependency on instructor AI proficiency and assumptions about institutional technology infrastructure create potential implementation barriers requiring careful consideration during adoption planning. Future research should examine adaptation strategies for diverse institutional contexts and identify critical success factors for effective implementation.

For the research community, priorities include empirical validation through systematic evaluation, development of assessment tools for measuring AI/ML educational effectiveness, and comparative studies of different pedagogical

approaches. Longitudinal studies tracking graduate career success and professional contributions would validate educational investment and inform continuous improvement efforts.

Recommendations and Vision for Future Development

For educators and institutions, recommendations include adoption and adaptation of the framework to local contexts and student needs, investment in faculty development and technology infrastructure, and collaboration with industry partners for curriculum relevance and student opportunities. Implementation should begin with pilot programs allowing refinement before full-scale deployment.

The vision for AI/ML education includes movement toward more accessible, inclusive education serving diverse populations while maintaining professional relevance and academic quality. Integration of ethical reasoning as fundamental to technical competency should become standard practice, ensuring graduates understand both technical capabilities and societal implications of AI/ML systems.

Continuous adaptation to technological advancement while maintaining pedagogical soundness represents ongoing challenge requiring systematic approaches to curriculum evolution and faculty development. The framework provides foundation for addressing these challenges while preparing students for responsible participation in the AI-driven future through comprehensive technical and professional competency development.

This systematic course development approach addresses current research gaps while providing practical institutional guidance for developing effective educational programs that serve diverse student populations and prepare graduates for rapidly evolving career demands in responsible AI development and deployment.

REFERENCES

- Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom*. AEHE-ERIC Higher Education Report. George Washington University.
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278. <https://doi.org/10.1109/ACCESS.2020.2988510>

- Clayton, P. R., & Clopton, J. (2019). Business curriculum redesign: Integrating data analytics. *Journal of Education for Business*, 94(1), 57-63. <https://doi.org/10.1080/08832323.2018.1502142>
- Coursera. (2024). Machine learning specialization [Online course]. Stanford University and DeepLearning.AI. <https://www.coursera.org/specializations/machine-learning-introduction>
- Dean, M. D. (2020). Using the learning assistant model in an undergraduate business analytics course. *INFORMS Transactions on Education*, 20(3), 125-133. <https://doi.org/10.1287/ited.2019.0221>
- Felder, R. M., & Brent, R. (2009). Active learning: An introduction. *ASQ Higher Education Brief*, 2(4), 1-5.
- Hooda, M. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2022/5215722>
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- Liu, X., & Burns, A. C. (2018). Designing a marketing analytics course for the digital age. *Marketing Education Review*, 28(1), 28-40. <https://doi.org/10.1080/10528008.2017.1421049>
- McCarthy, M. (2010). Experiential learning theory: From theory to practice. *Journal of Business & Economics Research*, 8(5), 131-139. <https://doi.org/10.19030/jber.v8i5.725>
- Mew, L. (2015). Implementing cloud computing for educational applications. *Information Systems Education Journal*, 13(4), 22-31.
- Mew, L. (2020). Designing and implementing an undergraduate data analytics program for non-traditional students. *Information Systems Education Journal*, 18(3), 18-27.
- MIT Professional Education. (2024). Professional certificate program in machine learning & artificial intelligence. <https://professional.mit.edu/course-catalog/professional-certificate-program-machine-learning-artificial-intelligence-0>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223-231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Sharp Sight Labs. (2016, March 22). The real prerequisite for machine learning isn't math, it's data analysis [Blog post]. *R-bloggers*. <https://www.r-bloggers.com/the-real-prerequisite-for-machine-learning-isnt-math-its-data-analysis/>
- UC Berkeley Executive Education. (2024). Professional certificate in machine learning and artificial intelligence. <https://em-executive.berkeley.edu/professional-certificate-machine-learning-artificial-intelligence>
- University of Texas at Austin. (2024). AI & machine learning certificate program online. <https://onlineexeced.mcombs.utexas.edu/online-ai-machine-learning-course>
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.
- World Economic Forum. (2025). *Future of Jobs Report 2025*. <https://www.weforum.org/publications/the-future-of-jobs-report-2025/>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, Y. G., Dang, M. Y., & Albritton, M. D. (2020). A foundation course in business analytics: Design and implementation at two universities. *Journal of Information Systems Education*, 31(4), 244-259

APPENDIX A

Syllabus Information

ISYS 398U: Artificial Intelligence and Machine Learning for Enterprise Applications (3 Credit Hours)

Course Description: This course provides a holistic review of Artificial Intelligence and Machine Learning (AI/ML) for the practitioner. The objective is to give students the tools to facilitate working in AI/ML at an entry level. Key algorithms such as regression for sales forecasting and clustering for customer segmentation are discussed and practiced. Practical tools for the enterprise such as Python, scikit-learn, cloud platforms for model deployment are introduced, and enterprise applications for improving cybersecurity, enhancing customer experience, and ethics in corporate AI are demonstrated. Current trends in the enterprise such as generative AI for content creation and AI agents are discussed.

Course Learning Outcomes

Upon successful completion of this course, students will be able to:

- Identify and explain the fundamental concepts, algorithms, and techniques used in artificial intelligence and machine learning
- Apply and evaluate AI/ML algorithms using industry standard tools, appropriate metrics, and optimization techniques to solve specific problems
- Develop real-world AI applications that demonstrate technical proficiency for a professional portfolio, including effective use of generative AI
- Assess potential ethical issues in AI applications and recommend appropriate mitigation strategies

Course Prerequisites: None (prompt engineering skills developed early in course support independent problem-solving)

Assessment Strategy

- Portfolio Projects (60%): Four progressive laboratory assignments building comprehensive GitHub repository
- Discussion Participation (20%): Industry-specific application discussions requiring original insights
- Quizzes (20%): Time-limited assessments reinforcing lesson attendance and concept comprehension

Course Modules and Learning Objectives

Unit 1: Introduction to AI and Generative AI

- Defines artificial intelligence and its core components, differentiating between narrow AI and general AI systems
- Explores various generative AI tools and large language models, including their capabilities and limitations
- Covers ethical considerations, biases, and evaluation methods for AI-generated content

Unit 2: Prompt Engineering

- Applies advanced prompt engineering techniques to optimize model performance and output quality
- Structures effective prompts using a framework with essential components (context, instructions, output format, rules, examples)
- Implements data-driven testing approaches to systematically improve prompt effectiveness and design specialized prompts for different applications

Unit 3: Machine Learning Foundations

- Identifies when machine learning is appropriate for solving problems versus traditional programming approaches
- Compares supervised, unsupervised, and reinforcement learning approaches with their respective use cases
- Distinguishes between classroom and production ML environments, outlining requirements for reliable and scalable ML systems

Unit 4: ML Process

- Describes the complete machine learning workflow from data collection through deployment as an iterative cycle
- Covers data preparation techniques including cleaning, transformation, and exploration as foundation for ML projects
- Emphasizes feature engineering as often more impactful than algorithm choice, and the importance of training/validation/testing splits

Unit 5: Supervised Learning

- Distinguishes between classification and regression problems for appropriate approach selection
- Applies linear and tree-based algorithms to solve real-world prediction problems
- Evaluates model performance using appropriate metrics (precision, recall, MAE, RMSE, R-squared) aligned with business consequences

Unit 6: Unsupervised Learning

- Discovers patterns in unlabeled data through clustering techniques and dimensionality reduction
- Addresses common model problems like overfitting and underfitting through cross-validation, feature selection, and ensemble methods
- Transforms "black box" algorithms into transparent tools through model explainability techniques, building trust and supporting regulatory compliance

Unit 7: MLOps

- Bridges the gap between ML development and operations through reproducible pipelines and systematic deployment
- Establishes core MLOps capabilities including reusable environments, model registration, tracking lineage, and implementing monitoring
- Evolves MLOps maturity from manual processes to fully automated pipelines requiring cultural change and cross-functional collaboration

Unit 8: AI Automations

- Implements AI-powered automation workflows using no-code/low-code platforms like Make and N8n to streamline business processes and reduce manual tasks
- Connects AI models and services through API integrations, webhooks, and data transformations to create intelligent automation chains
- Designs error handling, conditional logic, and human-in-the-loop checkpoints to ensure reliable and responsible automated decision-making systems

Unit 9: Deep Learning

- Covers neural network architectures and multi-layer networks for complex pattern recognition
- Explores deep learning advantages including automatic feature learning and handling raw data
- Addresses implementation considerations for deep learning in enterprise contexts

Unit 10: Computer Vision and Time Series

- Applies computer vision for quality control, inventory management, healthcare, and security applications
- Implements time series analysis for forecasting, anomaly detection, and pattern recognition in business contexts
- Addresses ethical considerations including explainability, bias prevention, and privacy protection in deep learning applications

Unit 11: Natural Language Processing (NLP)

- Defines NLP fundamentals and explains key preprocessing techniques for text data
- Analyzes major NLP tasks including text classification, sentiment analysis, topic modeling, and machine translation
- Develops strategies for implementing NLP projects in organizations using modern large language models

Unit 12: Emerging Technologies 2025

- Distinguishes AI agents from traditional automation through autonomous decision-making and adaptive problem-solving capabilities
- Evaluates Model Context Protocol (MCP) for standardizing AI connections to external data sources
- Assesses multimodal AI processing text, images, audio, and video together, plus low-code/no-code platforms for democratizing AI development

Sample Laboratory Assignment

Laboratory 1: Prompt Engineering for Business Applications

Students select an industry-specific business problem relevant to their professional context and develop systematic prompt engineering solutions using ChatGPT or similar large language models. Requirements include:

- Problem statement and business context documentation
- Iterative prompt development with minimum 3 variations
- Comparative analysis of prompt effectiveness

- Creation of prompt library for future use
- GitHub repository with complete documentation
- Reflection on ethical considerations and limitations

Deliverables must demonstrate progressive refinement of prompts, systematic analysis of different approaches, and creation of documentation suitable for professional portfolio presentation. Students present findings in industry-standard format appropriate for stakeholder communication.