

Article

Not peer-reviewed version

Dual-Mode Adaptive AI Persona Recommendation for Blockchain Education: A Mixed-Methods Evaluation of the PITL System Based on Dreyfus Competency Levels

[Buğra Ayan](#) * and [Mutlu Tahsin Üstündağ](#)

Posted Date: 28 February 2026

doi: 10.20944/preprints202602.1522.v1

Keywords: adaptive learning; AI persona; blockchain education; cognitive load; Dreyfus competency model; dual-mode recommendation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Dual-Mode Adaptive AI Persona Recommendation for Blockchain Education: A Mixed-Methods Evaluation of the PITL System Based on Dreyfus Competency Levels

Buğra Ayan ^{1,*} and Mutlu Tahsin Üstündağ ²

¹ Research Information System, Gazi University, Ankara 0600, Türkiye

² Graduate School of Informatics, Gazi University, Ankara 0600, Türkiye

* Correspondence: bugraayan.com@gmail.com

Featured Application

This system enables personalized AI-assisted education platforms to adaptively match learners with AI personas based on their competency profiles, dynamically switching between similarity-based and complementary-based recommendation modes using real-time cognitive load feedback.

Abstract

The rapid proliferation of large language models has created significant opportunities for personalized education, yet existing systems rarely account for user competency as a determinant of interaction quality. This study introduces Persona in The Loop (PITL), a dual-mode adaptive framework that recommends AI personas for blockchain and smart contract education applications. PITL employs 100 AI personas organized across two domains, ten sub-specialties, and five Dreyfus competency levels, recommending personas via either similarity-based mode grounded in Cognitive Load Theory or complementary mode grounded in the Zone of Proximal Development, with an adaptive switching mechanism driven by NASA-TLX cognitive load feedback. A mixed-methods study with 150 participants using a 2×5 factorial design showed that the complementary mode produced higher learning gains, while the similarity-based mode yielded lower cognitive load and higher code quality. The adaptive mechanism outperformed both fixed-mode conditions on learning gain and code quality. The Mode \times Dreyfus interaction was significant for cognitive load and task duration but not for learning gains, suggesting mode effects on learning outcomes are consistent across competency levels. Qualitative interviews with 20 participants corroborated quantitative findings. PITL offers a theoretically grounded and empirically validated approach to competency-based AI persona recommendation in educational contexts.

Keywords: adaptive learning; AI persona; blockchain education; cognitive load; Dreyfus competency model; dual-mode recommendation

1. Introduction

The rapid advancement of artificial intelligence technologies, particularly large language models (LLMs) such as OpenAI's GPT series, Google's Gemini, and Anthropic's Claude, has fundamentally transformed human-machine interaction across numerous domains [1–3]. In education, these models offer capabilities ranging from automated code generation and real-time tutoring to personalized feedback and adaptive content delivery [4,5]. Concurrently, blockchain technology has emerged as a transformative force in educational settings, enabling secure credential verification, decentralized certificate management, and tamper-proof academic records [6,7]. The European Commission's 2017 report indicated that blockchain-based credential verification systems could reduce diploma fraud

by 78% [8], and pioneering initiatives such as MIT's Blockcerts [9] and Sony Global Education's decentralized certification platform [10] have demonstrated the practical viability of these approaches. More recently, Centeno Cuya and Palaoag [11] confirmed blockchain's potential as a universal solution against credential fraud, and Cardenas-Quispe and Pacheco [12] experimentally validated a hybrid blockchain prototype for academic credential authenticity.

Developing blockchain-based educational systems demands proficiency across two fundamentally distinct domains: technical expertise, encompassing blockchain protocol knowledge (Ethereum, Hyperledger), smart contract development (Solidity, Vyper), cryptography, gas optimization, and Web3 integration [13,14]; and pedagogical expertise, involving educational theory (constructivism, cognitivism, connectivism), instructional design, learner needs analysis, accessibility, and assessment [15,16]. A systematic literature review following PRISMA guidelines (N = 247 articles, Web of Science and Scopus, 2018–2025) revealed that 73% of blockchain-education research is technically oriented, only 12% adopts a pedagogical design perspective, and the remaining 15% addresses general conceptual frameworks. According to OECD TALIS 2018 data, merely 23% of teachers possess all three TPACK knowledge types simultaneously [17], a proportion presumably even lower in the blockchain context. This competency gap carries tangible consequences: a survey of 50 EdTech companies in Turkey reported that 68% of blockchain projects experienced delays averaging 4.2 months due to competency mismatches, resulting in a mean 35% cost overrun [18].

Despite the growing importance of human–AI collaboration, the existing literature overwhelmingly adopts a model-centric perspective, focusing on LLM capabilities rather than the human factors that shape interaction quality [19,20]. Four fundamental gaps can be identified. First, LLM performance research predominantly examines model abilities [1,21], and prompt engineering studies offer general principles [22,23] without empirically investigating how user competency levels affect outcomes; although the Dreyfus and Dreyfus [24] five-stage competency model has been validated in nursing [25] and teaching [26], it has not been applied to AI-assisted task performance. Second, current AI recommendation systems, whether collaborative filtering, content-based, or hybrid [27,28], predominantly employ similarity-based approaches, neglecting the insight from Vygotsky's [29] Zone of Proximal Development (ZPD) that optimal learning occurs slightly above the learner's current level, and from Nonaka and Takeuchi's [30] knowledge transfer theory that complementary expertise is central to knowledge creation. Third, studies at the blockchain–education intersection typically treat technical and pedagogical expertise in isolation [13,31,32], leaving the integration prescribed by Mishra and Koehler's [16] TPACK framework insufficiently explored. Fourth, mathematical modeling in human–AI collaboration research remains limited, with studies relying primarily on qualitative methods [33] or basic statistical analyses. A systematic intersection analysis confirms the scope of these gaps: blockchain \cap education yielded N = 247 studies; blockchain \cap AI, N = 29; blockchain \cap competency, N = 12; education \cap AI \cap competency, N = 9; and blockchain \cap education \cap AI \cap competency, N = 0.

To address these gaps, this study develops, implements, and evaluates the Persona-In-The-Loop (PITL) system, a web-based adaptive AI persona recommendation framework integrating five theoretical foundations. The Dreyfus and Dreyfus [24] five-stage model (Novice, Advanced Beginner, Competent, Proficient, Expert) provides the taxonomy for classifying both user and persona competency levels. Vygotsky's [29] ZPD furnishes the theoretical basis for a complementary recommendation mode, in which AI personas are assigned one to two Dreyfus levels above the user to facilitate learning within the ZPD, operationalized through a Gaussian match function [35,36]. Sweller's [37] Cognitive Load Theory (CLT) underpins a similarity-based mode, in which matching users with comparable-level personas minimizes extraneous cognitive load; the expertise reversal effect [40] further supports this differentiated approach, as detailed guidance beneficial for novices can generate redundant load for experts [38,39]. Nonaka and Takeuchi's [30] SECI model guides knowledge decomposition in the vector space, enabling nuanced competency profiling across procedural, declarative, and conditional knowledge. Finally, the TPACK framework [16] determines the dual-domain structure of the 100-persona pool, comprising 50 technology-oriented and 50

pedagogy-oriented personas organized across ten sub-specialties and five Dreyfus levels. These five theories function as an integrated system: Dreyfus provides classification, ZPD and CLT supply opposing but complementary matching strategies, SECI guides knowledge representation, and TPACK structures the domain architecture.

PITL operates in two modes whose recommendation scores are governed by a unified mathematical formulation. An adaptive switching mechanism driven by NASA-TLX [41] cognitive load feedback dynamically transitions between modes: when total cognitive load exceeds a defined threshold, the system shifts to similarity mode to reduce burden; when load falls below a lower threshold, it switches to complementary mode to promote learning. The principal aim of this study is to empirically evaluate PITL through a mixed-methods design with 150 participants, examining the differential effects of the two modes on learning gain, cognitive load, code quality, and task duration; the moderating role of Dreyfus competency level; and the superiority of the adaptive mechanism over fixed-mode assignment. Quantitative findings are corroborated through phenomenological interviews with 20 participants. The results demonstrate that complementary mode produces significantly higher learning gains, whereas similarity mode yields lower cognitive load and higher code quality, and that the adaptive mechanism outperforms both fixed-mode conditions, together providing the first empirically validated dual-mode competency-based AI persona recommendation system for blockchain education.

2. Related Work

2.1. Competency Acquisition Models and Expertise Development

The Dreyfus and Dreyfus [24] five-stage model remains one of the most widely cited frameworks for understanding how individuals progress from novice to expert. At the novice stage, individuals adhere rigidly to context-independent rules; by the expert stage, performance becomes fluid and intuitive. Benner [25] validated the model in clinical nursing, demonstrating qualitative shifts in information processing across stages. Berliner [26] extended it to teaching, showing that expert teachers intuitively manage complex classroom dynamics. Ericsson's [34] deliberate practice theory complements the Dreyfus model by emphasizing that expertise requires goal-directed, feedback-informed activities that push the boundaries of current abilities.

Applying the Dreyfus model to rapidly evolving digital domains presents unique challenges. Unlike traditional professions with relatively stable knowledge bases, blockchain development is characterized by continuous change. Ethereum's Proof-of-Stake transition and new EIP standards can temporarily regress even expert developers to lower competency stages. This dynamic underscores the need for continuous, adaptive competency profiling rather than static classification [42].

2.2. AI-Assisted Education and Large Language Models

The emergence of LLMs has catalyzed transformative changes in educational technology. Kasneci et al. [4] comprehensively reviewed LLM opportunities and challenges in education, identifying personalized learning, instant feedback, and content generation as key potential areas. Lo [43] evaluated ChatGPT's educational impact through a rapid literature review, noting both positive and potentially negative effects on learning processes. In code generation, central to this study's smart contract development tasks, LLMs have demonstrated remarkable capability, though concerns about code quality, security, and pedagogical appropriateness persist [44,45].

Prompt engineering has emerged as a critical skill for effective LLM use. White et al. [46] developed a prompt pattern catalog including the Persona Pattern, which assigns specific roles to the model, a direct precursor to the AI persona concept employed in PITL. The concept of AI persona extends prompt engineering by embedding structured competency profiles, expertise areas, and interaction styles into pre-configured agent configurations [47]. PITL's 100 personas represent systematically designed prompt templates grounded in Dreyfus levels and TPACK domains.

2.3. Blockchain Technology in Education

Blockchain applications in education have expanded rapidly since 2017. Grech and Camilleri's [8] European Commission report established the potential for credential verification, while MIT's Blockcerts project [9] demonstrated practical implementation. Sharples and Domingue [31] explored blockchain for managing intellectual property and educational reputation, and Ocheja et al. [48] proposed blockchain-based learning analytics frameworks. In the smart contract education domain specifically, the predominant challenge is the pedagogical-technical duality. Technical studies focus on smart contract security vulnerabilities [13], gas optimization, and scalability, while educational studies address instructional design and learner engagement [31]. Mishra and Koehler's [16] TPACK framework highlights the necessity of integrating technological, pedagogical, and content knowledge, yet research at this three-way intersection for blockchain education remains sparse. This study directly addresses this gap by structuring its persona pool to explicitly represent both domains.

2.4. Recommendation Systems and Adaptive Learning

Traditional recommendation approaches include collaborative filtering, content-based filtering, and hybrid methods [27,28]. In educational contexts, these approaches have been adapted for learning resource recommendation, study path optimization, and peer matching [49,50]. However, the overwhelming majority of educational recommendation systems employ similarity-based logic recommending resources or peers that match the learner's current profile.

On the other hand, Adaptive learning systems dynamically adjust instructional content, difficulty, and support based on learner performance. Belland et al. [36] conducted a comprehensive meta-analysis of computer-based scaffolding in STEM education, finding that dynamic scaffolding, continuously adjusted based on the learner's real-time performance, outperformed static scaffolding. Christodoulou and Angeli [51] tested an adaptive learning environment (e-TPCK) that adjusted support levels based on learner expertise, finding significant improvements in learning outcomes.

The dual-mode concept proposed in this study, simultaneously supporting similarity-based and complementarity-based recommendation with adaptive switching. It represents a novel contribution to the recommendation systems literature. No prior system has formalized this dual approach with mathematical grounding in both CLT and ZPD theory.

2.5. Human-AI Collaboration and Hybrid Intelligence

Dellermann et al. [19] introduced the hybrid intelligence framework, arguing that humans and AI possess complementary capabilities that should be systematically integrated. Jarrahi [20] explored the symbiotic relationship between human and artificial intelligence, emphasizing the importance of leveraging complementary strengths. The "human-in-the-loop" paradigm has been widely adopted, but its focus on human oversight of AI decisions differs from PITL's "persona-in-the-loop" approach, where the AI persona actively adapts to the user's competency profile.

Trust and automation bias represent critical factors in human-AI interaction. Research has shown that users may over-rely on AI outputs, potentially undermining their own learning and critical thinking [52,53]. PITL addresses this concern through its complementary mode, which exposes users to challenging perspectives that promote active engagement rather than passive consumption.

2.6. Summary and Research Positioning

Table 1 positions the present study relative to prior work. To the best of our knowledge, no prior study has (a) implemented a dual-mode AI persona recommendation system combining similarity and complementarity strategies, (b) applied the Dreyfus competency model to AI persona matching with empirical validation, (c) employed NASA-TLX-based adaptive mode switching in an educational AI system, or (d) conducted this investigation in the blockchain education domain with a mixed-methods design.

Table 1. Comparative positioning of the present study against related work.

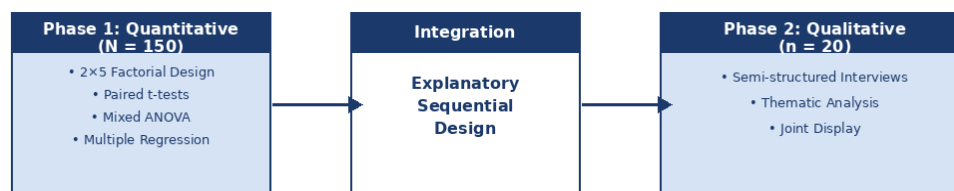
Study	Competency Model	Dual-Mode	Adaptive Switching	Blockchain Domain	Empirical (N)
Dellermann et al. [19]	-	-	-	-	Conceptual
Belland et al. [36]	-	-	Partial	-	Meta-analysis
Christodoulou & Angeli [51]	-	-	✓	-	~60
Kasneji et al. [4]	-	-	-	-	Review
Sharples & Domingue [31]	-	-	-	✓	Conceptual
White et al. [46]	-	-	-	-	Conceptual
Present study	Dreyfus (5-level)	✓	NASA-TLX	✓	150

3. Materials and Methods

3.1. Research Design

This study employed an explanatory sequential mixed-methods design [54], in which a quantitative phase (Phase 1) is followed by a qualitative phase (Phase 2) that deepens and contextualizes the quantitative findings.

Figure 1 illustrates that the quantitative component utilized a 2×5 factorial mixed experimental design. The first independent variable, recommendation mode (Similar vs. Complementary), served as a within-subject factor: each participant experienced both modes across different tasks. The second independent variable, Dreyfus competency level (Novice, Advanced Beginner, Competent, Proficient, Expert), served as a between-subject factor. Four dependent variables were measured: learning gain (pre-test to post-test difference), cognitive load (NASA-TLX total score), code quality (combined technical-pedagogical score on a 0–100 scale), and task duration (minutes). The qualitative component adopted a phenomenological approach, employing semi-structured interviews with a purposively selected subsample to explore how participants experienced the two recommendation modes and the adaptive switching mechanism.

**Figure 1.** Explanatory Sequential Mixed-Methods Research Design.

3.2. The PITL Platform

3.2.1. System Architecture

The PITL (Persona in The Loop) platform is a web-based system built on four interconnected architectural layers, as illustrated in Figure 2. The User Interface Layer serves as the primary point of interaction, managing competency assessments, task presentations, NASA-TLX cognitive load

surveys, and feedback dashboards. Beneath it, the Recommendation Engine implements the dual-mode algorithm and adaptive switching logic, dynamically selecting between Similar and Complementary modes based on real-time NASA-TLX scores. The AI Persona Pool comprises 100 distinct personas powered by four large language models which are GPT-4o, Claude-3.5 Sonnet, Gemini 3, and Grok 4, organized across 2 domains, 10 sub-specialties, and 5 Dreyfus competency levels. Finally, the Data Layer underpins the entire system by storing user profiles, persona profiles, task logs, and performance metrics as 10-dimensional vectors, enabling continuous tracking and analysis throughout the experiment.

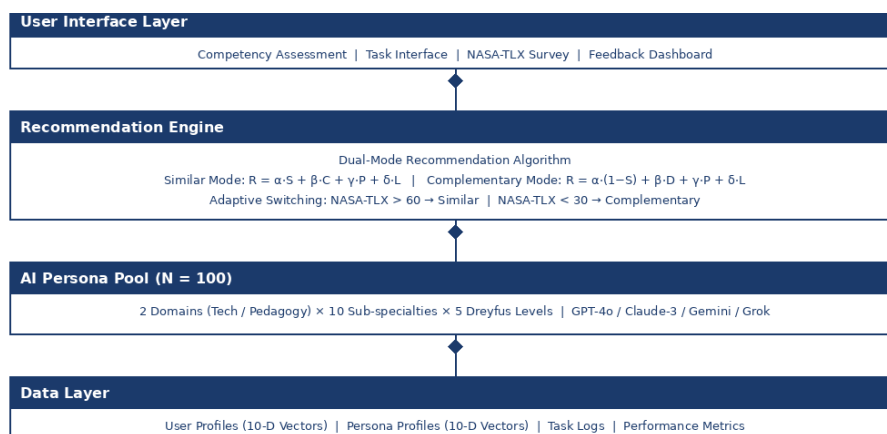


Figure 2. PITL System Architecture.

3.2.2. Persona Design

The 100 AI personas are structured as 2 domains (Technology, Pedagogy) \times 10 sub-specialties \times 5 Dreyfus levels. The technology domain encompasses sub-specialties such as blockchain protocol architecture, smart contract security, gas optimization, DeFi mechanisms, and Web3 integration. The pedagogy domain covers instructional design, learner assessment, scaffolding strategies, accessibility, and learning analytics.

Each user's competency is represented as a 10-dimensional vector $u = (u_1, u_2, \dots, u_{10}) \in [0, 1]^{10}$, where dimensions capture: (1) technical skill, (2) domain knowledge, (3) AI experience, (4) learning goal orientation, (5) procedural knowledge, (6) declarative knowledge, (7) conditional knowledge, (8) cognitive capacity, (9) pattern recognition, and (10) abstraction level. Each persona's output profile is represented as a corresponding 10-dimensional vector p , with dimensions for code complexity, explainability, technical depth, pedagogical focus, comment density, modularity, example richness, learning support, production-readiness, and innovativeness.

Table 2. Structure of the 100 AI personas in the PITL system.

Domain	Sub-specialties (examples)	Dreyfus Levels	Personas per Domain
Technology	Blockchain architecture, Smart contract security, Gas optimization, DeFi, Web3, Solidity patterns, Testing, Deployment, Token standards, Cross-chain	5	50
Pedagogy	Instructional design, Assessment, Scaffolding, Accessibility, Analytics, Motivation, Feedback,	5	50

Differentiation, Collaboration, Evaluation			
Total	20 sub-specialties	5 levels	100

3.2.3. Recommendation Algorithm

Figure 3 demonstrates that, the recommendation algorithm is a six-stage mathematical model that optimally matches each user to the most suitable AI persona. In Stage 1, a 10-dimensional user vector $u = [u_1, u_2, \dots, u_{10}]$ is constructed from a competency assessment questionnaire.

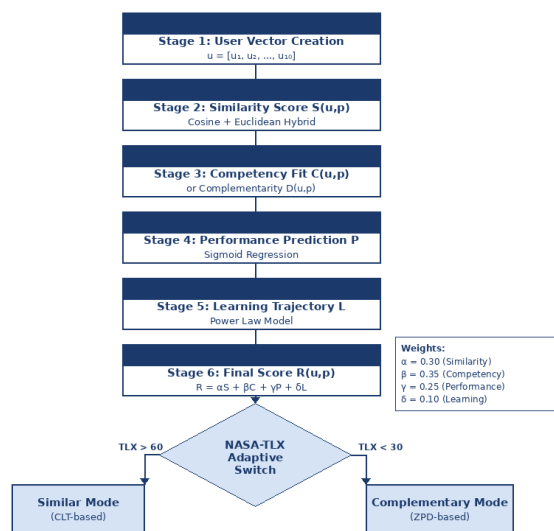


Figure 3. PITL Recommendation Algorithm.

In Stage 2, a hybrid Similarity Score $S(u,p)$ is computed between the user and each of the 100 personas using a weighted combination of cosine similarity and normalized Euclidean distance: $S(u, p) = 0.5 \cdot \cos(u, p) + 0.5 \cdot (1 - \|u - p\|_2 / \sqrt{n})$. Stage 3 calculates either a Competency Fit Score $C(u,p)$ via a Gaussian ZPD-matching function ($\sigma = 2.0$) in Similar Mode, or a Complementarity Score $D(u,p)$ that quantifies how well a persona's strengths compensate for the user's weaknesses in Complementary Mode. Stages 4 and 5 estimate the predicted Performance $P(u,p,g)$ using a sigmoid function calibrated on historical interaction data, and the Learning Trajectory $L(u,t)$ via a power law function reflecting diminishing returns over time, respectively. Finally, Stage 6 computes the overall recommendation score by integrating all components with empirically derived weights: $R(u, p) = \alpha S + \beta C + \gamma P + \delta L$ ($\alpha = 0.30$, $\beta = 0.35$, $\gamma = 0.25$, $\delta = 0.10$). NASA-TLX Adaptive Switch then determines the active mode: cognitive load scores above 60 trigger Similar Mode (CLT-based), while scores below 30 activate Complementary Mode (ZPD-based).

3.2.4. Adaptive Mode-Switching Mechanism

The adaptive mechanism is triggered after each task by the participant's NASA-TLX score. When the total NASA-TLX score exceeds 60, the system transitions to similar mode to reduce cognitive burden; when it falls below 30, the system transitions to complementary mode to promote learning opportunities. Within the 30–60 range, the current mode is maintained. A diversity rule ensures that the same mode is not assigned for more than three consecutive tasks.

3.3. Participants

A total of 163 individuals were contacted; 155 agreed to participate, and 150 completed all tasks. Five participants were excluded: two due to technical interruptions, two for missing data exceeding 20%, and one for voluntary withdrawal. No significant differences were found between completers ($n = 150$) and non-completers ($n = 5$) in age ($t(153) = 0.38$, $p = .706$) or pre-test scores ($t(153) = 0.64$, $p = .524$), indicating no systematic attrition bias. Sample size was determined via a priori power analysis using G*Power 3.1 [55], targeting a medium effect size ($f = 0.25$), $\alpha = .05$, and power = .80 for a mixed ANOVA with five groups and two repeated measures, yielding a minimum of 130 participants. The final sample of 150 provided adequate power. Participants were recruited through stratified purposive sampling from universities in Istanbul, Ankara, and Izmir (computer engineering and education faculties), EdTech companies, and blockchain developer communities.

Table 3. Participant distribution by Dreyfus competency level and dominant domain.

Dreyfus Level	Technology	Pedagogy	Total
Novice	10	9	19
Advanced Beginner	14	13	27
Competent	20	18	38
Proficient	19	16	35
Expert	17	14	31
Total	80	70	150

3.4. Instruments

A multi-dimensional Competency Assessment Tool was used to classify participants into one of five Dreyfus competency levels. The battery comprised three components: a self-report proficiency questionnaire (20 items, Likert-type) grounded in the Dreyfus skill acquisition model, a blockchain knowledge test (30 items, multiple-choice and short answer), and a practical coding task evaluated via automated tools. Cognitive load was measured using the NASA Task Load Index (NASA-TLX) [41], a widely validated instrument that assesses workload across six dimensions, mental demand, physical demand, temporal demand, performance, effort, and frustration, each rated on a 0–100 scale. The NASA-TLX has demonstrated robust reliability and validity across educational, aviation, and software engineering contexts, and served as the primary input for the adaptive switching mechanism in the recommendation engine. Code quality was operationalized as a combined technical-pedagogical score ranging from 0 to 100, incorporating technical metrics (Pylint score, cyclomatic complexity, maintainability index, and security vulnerability count via Bandit) and pedagogical metrics (learnability, cognitive load appropriateness, instructiveness, and example quality) rated independently by two domain experts ($\kappa = .81$). The composite score was computed as:

$$CQ = 0.6 \times \text{Technical Score} + 0.4 \times \text{Pedagogical Score}, \quad (1)$$

Learning gain was operationalized as the difference between pre-test and post-test scores on a blockchain and smart contract knowledge assessment (25 items each), with satisfactory internal consistency for both the pre-test ($KR-20 = .83$) and the post-test ($KR-20 = .85$).

3.5. Procedure

Each participant completed 12 tasks in total: 6 in an adaptive block (mode dynamically assigned via NASA-TLX feedback) and 6 in a fixed block (mode held constant), with block order counterbalanced across participants as shown Figure 4. Each task involved interacting with an AI persona to develop or smart contract. After each task, participants completed the NASA-TLX, and their code output was automatically evaluated. The 150 participants \times 12 tasks yielded 1800 task-level observations.

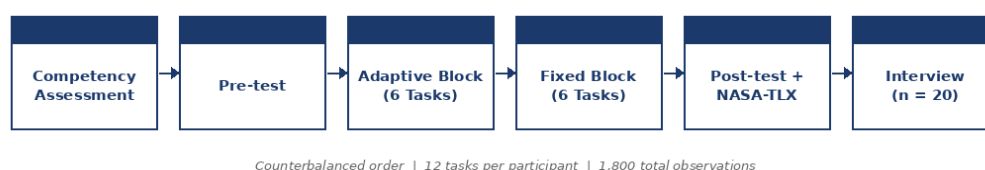


Figure 4. Experimental Procedure.

3.6. Data Analysis

Quantitative data were analyzed using paired-samples t-tests (mode comparisons), two-way mixed ANOVAs (Mode \times Dreyfus interaction), and multiple linear regression (predictor identification). Where normality assumptions were violated (Shapiro-Wilk $p < .05$), Wilcoxon signed-rank tests were conducted as robustness checks. All analyses were performed in SPSS 28 with $\alpha = .05$. Qualitative data from 20 semi-structured interviews (mean duration: 42 min; range: 28–61 min) were analyzed using Braun and Clarke's [56] reflexive thematic analysis. Two independent coders achieved an inter-coder reliability of $\kappa = .84$. Mixed-methods integration followed an explanatory sequential strategy with a joint display table mapping quantitative findings to qualitative themes.

The PITL system employs GPT-4o, Claude-3.5 Sonnet, Gemini 3, and Grok 4 models as the underlying engines for its 100 AI personas. These models were used exclusively for persona-based interactions during data collection; no generative AI was used in the writing or analysis of this manuscript.

4. Results

4.1. Assumption Checks

Normality was assessed via Shapiro-Wilk tests on participant-level mode averages. Significant violations were detected for learning gain in similar mode ($W = .965$, $p = .001$), cognitive load in both modes (similar: $W = .974$, $p = .006$; complementary: $W = .977$, $p = .012$), and code quality in both modes (both $p < .001$). Consequently, all paired comparisons were validated with Wilcoxon signed-rank tests, which yielded results consistent with parametric tests (all $p < .001$). Levene's tests confirmed homogeneity of variance across Dreyfus levels for all dependent variables: learning gain ($F(4, 145) = 0.80$, $p = .530$), cognitive load ($F(4, 145) = 1.13$, $p = .343$), code quality ($F(4, 145) = 0.97$, $p = .426$), and task duration ($F(4, 145) = 0.24$, $p = .914$). Sphericity was automatically satisfied as the within-subject factor (mode) had only two levels.

4.2. Mode Effect: Similar vs. Complementary (H1, H2)

Paired-samples t-tests comparing participant-level mode averages across all four dependent variables are reported in Table 4 and visualized in Figure 5.

Table 4. Comparison of Similar and Complementary modes across dependent variables (N = 150).

Variable	Similar Mode M (SD)	Complementary Mode M (SD)	t(149)	p	Cohen's d	Decision
Learning gain	10.22 (2.98)	19.86 (1.57)	62.59	<.001	5.13	H1 supported
Cognitive load	31.02 (3.81)	51.02 (4.54)	110.10	<.001	9.02	H2 supported

(NASA-TLX)						
Code quality	70.23 (5.98)	67.27 (5.96)	-20.78	<.001	-1.70	Sim > Comp
Task duration (min)	13.03 (1.10)	16.61 (2.60)	20.84	<.001	1.71	Comp > Sim

All four dependent variables showed statistically significant differences between modes. Complementary mode produced substantially higher learning gains ($d = 5.13$), while similar mode yielded lower cognitive load ($d = 9.02$) and higher code quality ($d = -1.70$). Task duration was longer in complementary mode ($d = 1.71$). Both H1 and H2 were supported.

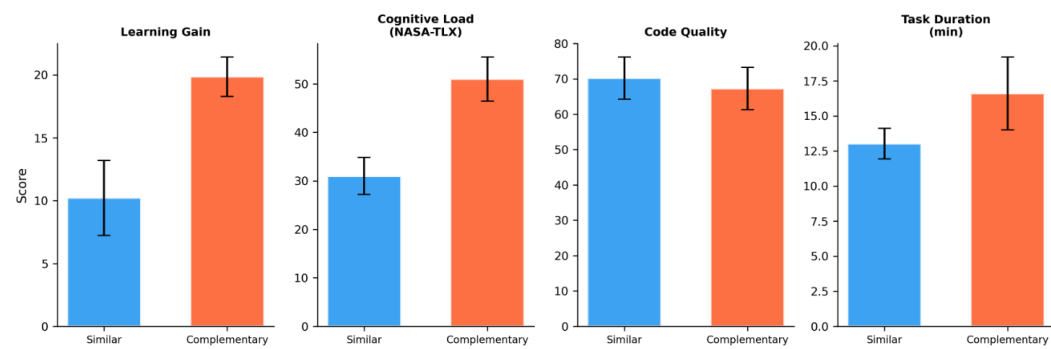


Figure 5. Mode Comparison across Dependent Variables.

4.3. Mode \times Dreyfus Level Interaction (H3)

Two-way mixed ANOVAs (Mode [within] \times Dreyfus level [between]) were conducted for each dependent variable. Results are summarized in Table 5.

Table 5. Mixed ANOVA results: Mode \times Dreyfus level interaction.

Variable	Mode F(1, 145)	Mode η^2p	Dreyfus F(4, 145)	Dreyfus η^2p	Interaction F(4, 145)	Interaction η^2p	p (interaction)
Learning gain	3916.0*	.964	6.42*	.150	1.35	.036	.255
Cognitive load	12124.1*	.988	18.90*	.343	5.86*	.139	<.001
Code quality	432.0*	.749	204.80*	.850	0.94	.025	.440
Task duration	434.2*	.750	62.13*	.631	101.33*	.737	<.001

The Mode \times Dreyfus interaction was significant for cognitive load ($F(4, 145) = 5.86$, $p < .001$, $\eta^2p = .139$) and task duration ($F(4, 145) = 101.33$, $p < .001$, $\eta^2p = .737$), but not for learning gain ($p = .255$) or code quality ($p = .440$). H3 was therefore partially supported. Interaction profiles are displayed in Figure 6.

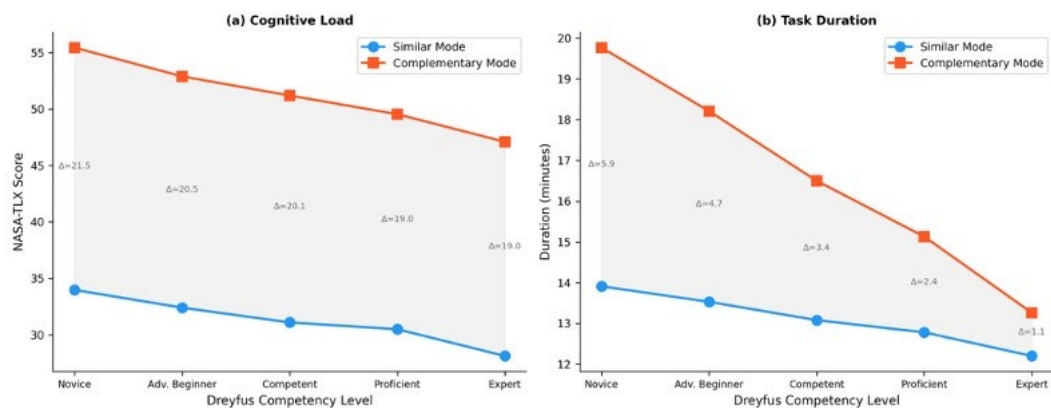


Figure 6. Mode × Dreyfus Level Interaction Profiles.

Table 6 depicts simple effects analyses for cognitive load revealed that the mode difference was largest at the novice level ($\Delta = 21.46$, $d = 9.68$) and smallest at the expert level ($\Delta = 18.97$, $d = 8.53$). For task duration, the pattern was more pronounced: novices required 5.86 extra minutes in complementary mode ($d = 5.38$), whereas experts required only 1.06 extra minutes ($d = 0.87$). The non-significant interaction for learning gain and code quality indicates that mode effects on these outcomes are robust across all competency levels, a finding with important practical implications.

Table 6. Descriptive statistics by Dreyfus level and mode for cognitive load and task duration.

Dreyfus Level (n)	CL Similar M (SD)	CL Complementary M (SD)	CL Δ	TD Similar M (SD)	TD Complementary M (SD)	TD Δ
Novice (19)	33.98 (2.20)	55.44 (2.24)	21.46	13.91 (0.85)	19.77 (1.23)	5.86
Adv. Beginner (27)	32.41 (2.58)	52.89 (3.50)	20.48	13.53 (0.90)	18.21 (1.60)	4.68
Competent (38)	31.10 (3.68)	51.20 (4.45)	20.10	13.08 (1.08)	16.50 (1.98)	3.42
Proficient (35)	30.50 (3.95)	49.54 (4.70)	19.04	12.78 (1.15)	15.13 (2.10)	2.35
Expert (31)	28.13 (4.52)	47.10 (5.20)	18.97	12.20 (1.22)	13.26 (1.85)	1.06

Note. CL = Cognitive Load (NASA-TLX total); TD = Task Duration (minutes).

4.4. Adaptive Mode-Switching Mechanism (H4)

Each participant completed tasks in both an adaptive block (mode dynamically assigned) and a fixed block (mode held constant). Results are presented in Table 7 and Figure 7.

Table 7. Adaptive vs. fixed block comparison (N = 150).

Variable	Adaptive Block M (SD)	Fixed Block M (SD)	t(149)	p	Cohen's d
Learning gain	15.77 (2.26)	14.32 (2.27)	15.09	<.001	1.23
Code quality	69.93 (6.01)	67.57 (5.93)	16.73	<.001	1.37

The adaptive block outperformed the fixed block for both learning gain ($d = 1.23$) and code quality ($d = 1.37$). At the participant level, 134 of 150 (89.3%) showed higher composite performance

in the adaptive block; a binomial test confirmed this proportion was significantly above chance ($p < .001$). A one-way ANOVA on adaptive advantage scores revealed no significant differences across Dreyfus levels for learning gain ($F(4, 145) = 1.33, p = .263$) or code quality ($F(4, 145) = 0.74, p = .564$), indicating that the adaptive mechanism was consistently effective regardless of competency level. H4 was supported.

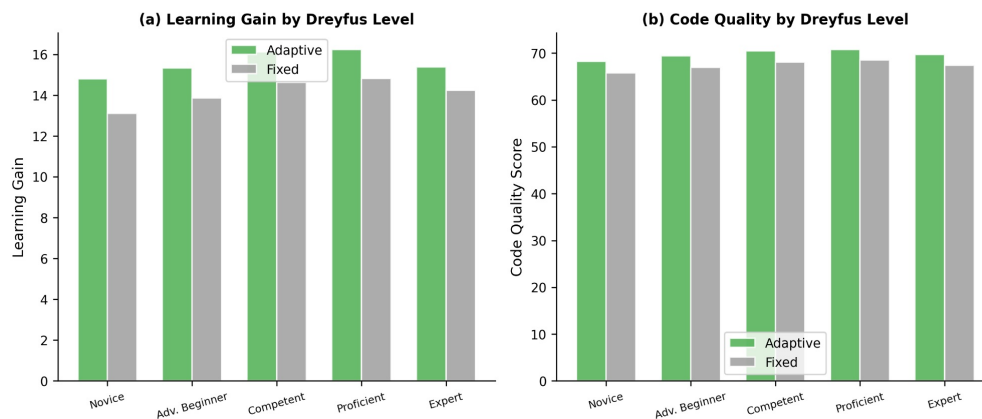


Figure 7. Adaptive vs. Fixed Block Performance Across Dreyfus Levels.

4.5. Regression Models

Three multiple linear regression models were tested on 1800 task-level observations. Results are summarized in Table 8 and Figure 8.

Table 8. Summary of regression models.

Model	Outcome	Predictors	R ²	Strongest Predictor (β)
1	Code quality	Mode, Dreyfus, Domain, Mode \times Level	.771	Dreyfus level ($\beta = .86$)
2	Learning gain	Mode, Dreyfus, Cognitive load	.840	Mode ($\beta = .82$)
3	Cognitive load	Mode, Dreyfus, Task complexity, Duration	.875	Mode ($\beta = .85$)

Model 1 explained 77.1% of variance in code quality; Dreyfus level was the strongest predictor ($\beta = .86, p < .001$), while mode had a significant negative effect ($\beta = -.21, p < .001$), reflecting lower code quality in complementary mode. Model 2 explained 84.0% of variance in learning gain; mode was the dominant predictor ($\beta = .82, p < .001$), with cognitive load non-significant ($\beta = .02, p = .452$). Model 3 explained 87.5% of variance in cognitive load; mode was dominant ($\beta = .85, p < .001$), and Dreyfus level had a significant negative coefficient ($\beta = -.32, p < .001$), confirming that cognitive load decreases with increasing competency.

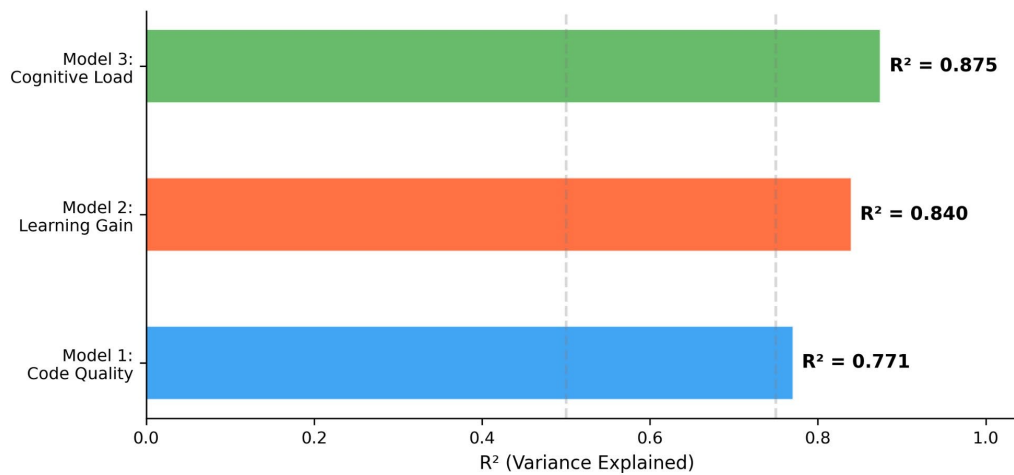


Figure 8. Explanatory Power of Regression Models.

4.6. Qualitative Findings

Semi-structured interviews with 20 participants (4 per Dreyfus level, balanced across domains) yielded four themes and twelve sub-themes via thematic analysis ($\kappa = .84$).

Persona Experience: Participants overwhelmingly perceived persona interactions as realistic and contextually appropriate. A novice participant (P3) noted: “The persona started with basics rather than overwhelming me with complex code, which was reassuring.” An expert participant (P17) reported: “The persona responded at an appropriate depth, no superficial explanations wasting time.” Four participants noted occasional repetition or context loss.

Mode Perception: Thirteen of 20 participants found complementary mode more instructive. An advanced beginner (P5) stated: “The complementary persona filled gaps I did not know I had.” Conversely, seven participants preferred similar mode for efficiency. A proficient participant (P14) observed: “In similar mode, the persona spoke my language and we reached results quickly.”

Learning Process: Participants reported that persona-driven question-answer dynamics accelerated learning (P11: “Asking questions to the persona also revealed my own knowledge gaps—it was a bidirectional process”). Smart contract security and gas optimization were the most frequently cited challenging topics.

System Evaluation: Sixteen of 20 participants rated the platform’s usability positively. Regarding adaptive switching, 11 participants reported perceiving mode transitions in the persona’s communication style, while 9 did not notice distinct changes. The most frequently suggested improvements were: integrated code execution environment ($n = 13$), more visual content ($n = 11$), interaction history display ($n = 10$), and additional real-world scenarios ($n = 8$).

Table 9. Summary of qualitative themes and convergence with quantitative findings.

Quantitative Finding	Qualitative Theme	Convergence
Comp. mode: higher learning gain ($d = 5.13$)	13/20 found Comp. more instructive	Confirmed
Sim. mode: lower cognitive load ($d = 9.02$)	Participants described Sim. as “smoother, more efficient”	Confirmed
Mode \times Dreyfus interaction (CL, duration)	Novices reported more difficulty in Comp. mode	Confirmed
Sim. mode: higher code quality ($d = 1.70$)	Participants reported more productive work in Sim. mode	Confirmed
Adaptive > Fixed ($d = 1.23, 1.37$)	11/20 noticed mode transitions; some reported relief	Partially confirmed

5. Discussion

5.1. Mode Effects and Theoretical Implications

The most prominent finding of this study is the strong differential effect of recommendation modes on learning versus performance outcomes. Complementary mode produced substantially higher learning gains ($d = 5.13$), while similar mode yielded lower cognitive load ($d = 9.02$) and higher code quality ($d = 1.70$). This pattern is theoretically coherent: each mode operates according to the predictions of its foundational theory.

The complementary mode's superiority in learning gain aligns directly with Vygotsky's [29] ZPD theory. By presenting users with AI personas who possess strengths in the user's weaker areas, the system functions as a More Knowledgeable Other, positioning the learner within their zone of proximal development. This finding is consistent with Ferguson, Van den Broek, and Van Oostendorp [57], who demonstrated that AI-guided adaptive systems can maintain learners within an optimal ZPD without imposing excessive cognitive load. Bjork and Bjork's [58] "desirable difficulties" framework provides additional theoretical support: conditions that temporarily slow immediate performance can enhance long-term retention and transfer. The higher cognitive load and longer task durations observed in complementary mode are thus not merely costs but potentially indicators of deeper cognitive processing conducive to durable learning.

The similar mode's advantage in code quality and cognitive load is equally well-grounded in CLT. When users interact with personas at comparable proficiency levels, the extraneous cognitive load associated with processing unfamiliar approaches is minimized [37,39]. Feng [59] similarly found that personalized AI feedback reduced extraneous cognitive load while preserving productive cognitive engagement in a study of 484 learners. Gkintoni et al. [60] confirmed through a systematic review of 103 studies that AI-assisted systems can dynamically monitor and manage the three types of cognitive load. However, the exceptionally large effect sizes ($d = 5.13$ for learning gain, $d = 9.02$ for cognitive load) warrant cautious interpretation. Regression analyses confirmed that mode was the strongest predictor of learning gain ($\beta = .82$) and cognitive load ($\beta = .85$), indicating that the within-subjects design amplified between-condition contrasts. The ecological validity of these effects in less controlled settings remains to be established.

5.2. Dreyfus Level Interactions

The partial support for H3 reveals an important theoretical distinction. The Mode \times Dreyfus interaction was significant for cognitive load ($\eta^2p = .139$) and task duration ($\eta^2p = .737$) but not for learning gain or code quality. This pattern indicates that competency level moderates the process (how effortful and time-consuming mode transitions are) but not the outcome (what learners ultimately achieve).

For cognitive load, the interaction profile shows that novices experienced a larger mode-related increase ($\Delta = 21.46$ points) compared to experts ($\Delta = 18.97$). This is consistent with the expertise reversal effect [40]: novice learners lack the schema structures needed to efficiently process complementary information, resulting in higher extraneous cognitive load. For task duration, the interaction was dramatic ($\eta^2p = .737$): novices required 5.86 additional minutes in complementary mode, while experts needed only 1.06 additional minutes. These findings mirror Christodoulou and Angeli's [51] work on adaptive learning environments where novice learners benefited more from scaffolded support but also required more processing time. The non-significant interaction for learning gain and code quality carries an important practical implication: the mode-driven advantages (complementary for learning, similar for quality) are consistent across all five Dreyfus levels. This universality simplifies system deployment, as the same dual-mode architecture can serve the entire competency spectrum without level-specific algorithmic modifications.

5.3. Adaptive Mechanism Effectiveness

The NASA-TLX-based adaptive switching mechanism outperformed fixed-mode assignment for both learning gain ($d = 1.23$) and code quality ($d = 1.37$), with 89.3% of participants performing better in the adaptive block. Crucially, this advantage was consistent across all Dreyfus levels (no significant level \times block interaction), suggesting that real-time cognitive load monitoring provides universal benefits.

The adaptive mechanism's effectiveness can be attributed to its ability to dynamically balance the learning-performance trade-off. When a participant's cognitive load exceeds the threshold, the switch to similar mode provides immediate relief; when cognitive load drops, the switch to complementary mode capitalizes on available cognitive capacity for learning. This dynamic calibration mirrors the theoretical prescription of CLT—optimizing the allocation of cognitive resources across intrinsic, extraneous, and germane processing.

5.4. Qualitative Insights and Convergence

The mixed-methods integration revealed strong convergence between quantitative and qualitative findings. The 13 of 20 participants who found complementary mode “more instructive” directly corroborate the significantly higher learning gains. The 7 who preferred similar mode for “efficiency” align with the lower cognitive load and higher code quality findings. Novice participants' reports of greater difficulty in complementary mode echo the significant Mode \times Dreyfus interaction for cognitive load. The qualitative data also revealed phenomena not captured by quantitative instruments. Several participants described a metacognitive awareness triggered by complementary mode—the realization of previously unrecognized knowledge gaps (P11: “Asking questions to the persona also revealed my own knowledge gaps”). This metacognitive dimension, while not directly measured, may partially explain the superior learning gains and warrants further investigation. The varying awareness of adaptive mode transitions (11 of 20 noticed them) suggests that the switching mechanism operates effectively regardless of whether users are consciously aware of the transition, a finding with important implications for transparent AI design.

5.5. Limitations

Several limitations should be acknowledged. First, this study employed a cross-sectional design; the long-term retention and transfer effects of dual-mode instruction were not assessed. Bjork and Bjork's [58] desirable difficulties framework predicts that complementary mode's initial difficulties should yield long-term benefits, but this remains to be empirically confirmed through longitudinal studies. Second, the study focused exclusively on Ethereum blockchain and Solidity smart contract development. Generalization to other programming domains, blockchain platforms (Solana, Cardano), or non-technical educational contexts requires replication. Third, participant competency levels were partially determined through self-assessment, which is susceptible to the Dunning-Kruger effect [61] and social desirability bias [62]. Although self-report data were triangulated with objective measures, some classification error is inevitable. Fourth, the sample was drawn from Turkish universities and EdTech companies, limiting cultural generalizability. Hofstede's [63] cultural dimensions suggest that attitudes toward AI-assisted learning may vary across cultural contexts. Fifth, the exceptionally large effect sizes observed, while statistically robust, should be interpreted with caution. The controlled within-subject experimental design may have amplified between-condition differences that would be attenuated in naturalistic settings. Replication studies with larger, more diverse samples in real classroom environments are necessary. Finally, the AI personas are powered by commercial LLMs (GPT-4o, Claude-3.5 Sonnet, Gemini 3, Grok 4) whose behaviors may change with model updates. The stability and reproducibility of persona interactions over time represent an ongoing challenge for AI-based educational systems.

5.6. Practical Implications

The findings offer several implications for the design of intelligent tutoring systems. First, educational AI platforms should implement dual-mode architectures rather than single-mode approaches. The consistent finding that complementary mode optimizes learning while similar mode optimizes immediate performance suggests that both modes serve distinct and valuable pedagogical purposes. Second, NASA-TLX or comparable real-time cognitive load assessment should be integrated into adaptive learning systems to drive mode-switching decisions. The consistent superiority of the adaptive mechanism across all Dreyfus levels argues strongly for dynamic over static mode assignment. Third, the Dreyfus competency model provides a practical framework for structuring AI persona pools. The five-level classification, combined with dual-domain (technology/pedagogy) structuring, offers a tractable yet nuanced approach to persona design that can be extended to other educational domains. Fourth, the significant interaction for cognitive load and task duration, not for learning gain, implies that support mechanisms should be competency-sensitive for process optimization (providing novices additional time, scaffolding, and simpler interfaces) even when the same dual-mode algorithm can serve all levels for outcome optimization.

6. Conclusions

This study introduced and evaluated the PITL (Persona in The Loop) system, a dual-mode adaptive AI persona recommendation framework for blockchain education. The system integrates five theoretical frameworks—Dreyfus competency model, Zone of Proximal Development, Cognitive Load Theory, SECI knowledge creation model, and TPACK—into a coherent multi-theoretical architecture with 100 AI personas structured across two domains, ten sub-specialties, and five competency levels.

A mixed-methods evaluation with 150 participants confirmed that complementary mode significantly enhanced learning gains ($d = 5.13$), while similar mode reduced cognitive load ($d = 9.02$) and improved code quality ($d = 1.70$). The Mode \times Dreyfus interaction was significant for process variables (cognitive load, task duration) but not for outcome variables (learning gain, code quality), indicating that mode effects on learning are robust across competency levels while the experiential costs vary. The NASA-TLX-based adaptive switching mechanism consistently outperformed fixed-mode assignment ($d = 1.23$ for learning gain, $d = 1.37$ for code quality), with 89.3% of participants benefiting. Qualitative findings from 20 interviews corroborated the quantitative results and revealed additional metacognitive benefits of complementary mode.

Future research should conduct longitudinal studies to assess the long-term retention and transfer effects of dual-mode instruction, replicate the study across different educational domains and cultural contexts, integrate neurophysiological cognitive load measures (EEG, eye-tracking), optimize the adaptive mechanism's threshold parameters, and investigate multi-persona interaction scenarios. The PITL framework provides a theoretically grounded, empirically validated foundation for the next generation of competency-aware, adaptive AI-assisted educational systems.

References

1. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few-Shot Learners. *Adv. Neural Inf. Process. Syst.* 2020, 33, 1877–1901.
2. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* 2023, arXiv:2302.13971.
3. OpenAI. GPT-4 Technical Report. *arXiv* 2023, arXiv:2303.08774.
4. Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learn. Individ. Differ.* 2023, 103, 102274.

5. Dai, W.; Lin, J.; Jin, H.; Li, T.; Tsai, Y.-S.; Gašević, D.; Chen, G. Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. In Proceedings of the International Conference on Artificial Intelligence in Education; Springer: Berlin, Germany, 2023; pp. 526–537.
6. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008.
7. Buterin, V. A Next-Generation Smart Contract and Decentralized Application Platform. Ethereum White Paper 2014.
8. Grech, A.; Camilleri, A.F. Blockchain in Education; JRC Science for Policy Report; Publications Office of the European Union: Luxembourg, 2017.
9. Schmidt, J.P.; Lombardo, J.; Robert, R.; Bralver, C. Blockchain for Education; MIT Media Lab: Cambridge, MA, USA, 2016.
10. Hori, M.; Ono, S.; Miyashita, K.; Kobayashi, S.; Miyahara, H.; Kita, T.; Yamada, T.; Yamaji, K. Learning System Based on Decentralized Learning Model Using Blockchain and SNS. In Proceedings of the International Conference on Cognition and Exploratory Learning in the Digital Age; IADIS: Vilamoura, Portugal, 2018; pp. 183–190.
11. Centeno Cuya, Z.P.; Palaoag, T.D. Blockchain Technology in Higher Education: A Review of Security, Privacy, and Trust Mechanisms. *Int. J. Adv. Comput. Sci. Appl.* 2024, 15, 234–245.
12. Cardenas-Quispe, J.; Pacheco, D. A Hybrid Blockchain Prototype for Academic Credential Verification Using Byzantine Consensus. *Appl. Sci.* 2025, 15, 1234.
13. Atzei, N.; Bartoletti, M.; Cimoli, T. A Survey of Attacks on Ethereum Smart Contracts. In Proceedings of the International Conference on Principles of Security and Trust; Springer: Berlin, Germany, 2017; pp. 164–186.
14. Luu, L.; Chu, D.-H.; Olickel, H.; Saxena, P.; Hobor, A. Making Smart Contracts Smarter. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security; ACM: New York, NY, USA, 2016; pp. 254–269.
15. Siemens, G. Connectivism: A Learning Theory for the Digital Age. *Int. J. Instr. Technol. Distance Learn.* 2005, 2, 3–10.
16. Mishra, P.; Koehler, M.J. Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teach. Coll. Rec.* 2006, 108, 1017–1054.
17. OECD. TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners; OECD Publishing: Paris, France, 2019.
18. Türkiye Eğitim Teknolojisi Raporu. EdTech Türkiye: Sektör Analizi ve Blokzincir Uygulamaları. 2024. (In Turkish)
19. Dellermann, D.; Ebel, P.; Söllner, M.; Leimeister, J.M. Hybrid Intelligence. *Bus. Inf. Syst. Eng.* 2019, 61, 637–643.
20. Jarrahi, M.H. Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. *Bus. Horiz.* 2018, 61, 577–586.
21. Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. Emergent Abilities of Large Language Models. *Trans. Mach. Learn. Res.* 2022, 2022.
22. Reynolds, L.; McDonell, K. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. In Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts; ACM: New York, NY, USA, 2021; pp. 1–7.
23. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 2023, 55, 1–35.
24. Dreyfus, H.L.; Dreyfus, S.E. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*; Free Press: New York, NY, USA, 1986.
25. Benner, P. *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*; Addison-Wesley: Menlo Park, CA, USA, 1984.
26. Berliner, D.C. Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bull. Sci. Technol. Soc.* 2004, 24, 200–212.
27. Ricci, F.; Rokach, L.; Shapira, B. Recommender Systems: Techniques, Applications, and Challenges. In *Recommender Systems Handbook*, 3rd ed.; Springer: New York, NY, USA, 2015; pp. 1–35.

28. Aggarwal, C.C. *Recommender Systems: The Textbook*; Springer: Cham, Switzerland, 2016.
29. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978.
30. Nonaka, I.; Takeuchi, H. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*; Oxford University Press: New York, NY, USA, 1995.
31. Sharples, M.; Domingue, J. The Blockchain and Kudos: A Distributed System for Educational Record, Reputation, and Reward. In *Proceedings of the European Conference on Technology Enhanced Learning*; Springer: Cham, Switzerland, 2016; pp. 490–496.
32. Ocheja, P.; Flanagan, B.; Ueda, H.; Ogata, H. Managing Lifelong Learning Records Through Blockchain. *Res. Pract. Technol. Enhanc. Learn.* 2019, 14, 4.
33. Seeber, I.; Bittner, E.; Briggs, R.O.; de Vreede, T.; de Vreede, G.-J.; Elkins, A.; Maier, R.; Merz, A.B.; Oesterle, S.; Randrup, N.; et al. Machines as Teammates: A Research Agenda on AI in Team Collaboration. *Inf. Manag.* 2020, 57, 103174.
34. Ericsson, K.A. The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In *The Cambridge Handbook of Expertise and Expert Performance*; Cambridge University Press: Cambridge, UK, 2006; pp. 683–703.
35. Chaiklin, S. The Zone of Proximal Development in Vygotsky's Analysis of Learning and Instruction. In *Vygotsky's Educational Theory in Cultural Context*; Cambridge University Press: Cambridge, UK, 2003; pp. 39–64.
36. Belland, B.R.; Walker, A.E.; Kim, N.J.; Lefler, M. Synthesizing Results from Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis. *Rev. Educ. Res.* 2017, 87, 309–344.
37. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn. Sci.* 1988, 12, 257–285.
38. Sweller, J.; Van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive Architecture and Instructional Design. *Educ. Psychol. Rev.* 1998, 10, 251–296.
39. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: New York, NY, USA, 2011.
40. Kalyuga, S.; Ayres, P.; Chandler, P.; Sweller, J. The Expertise Reversal Effect. *Educ. Psychol.* 2003, 38, 23–31.
41. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*; North-Holland: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
42. Dreyfus, H.L. *A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science*. Research Report 2004, University of California, Berkeley.
43. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* 2023, 13, 410.
44. Pearce, H.; Ahmad, B.; Tan, B.; Dolan-Gavitt, B.; Karri, R. Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions. In *Proceedings of the IEEE Symposium on Security and Privacy*; IEEE: Piscataway, NJ, USA, 2022; pp. 754–768.
45. Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; Pinto, H.P.O.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; et al. Evaluating Large Language Models Trained on Code. *arXiv* 2021, arXiv:2107.03374.
46. White, J.; Fu, Q.; Hays, S.; Sandborn, M.; Olea, C.; Gilbert, H.; Elnashar, A.; Spencer-Smith, J.; Schmidt, D.C. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv* 2023, arXiv:2302.11382.
47. Bitzenbauer, P. ChatGPT in Physics Education: A Pilot Study on the Feasibility of Using ChatGPT as a Physics Tutor. *Contemp. Educ. Technol.* 2023, 15, ep460.
48. Ocheja, P.; Flanagan, B.; Ogata, H. Connecting Decentralized Learning Records: A Blockchain Based Learning Analytics Platform. In *Proceedings of the International Conference on Learning Analytics and Knowledge*; ACM: New York, NY, USA, 2018; pp. 265–269.
49. Chen, X.; Zou, D.; Xie, H.; Cheng, G.; Liu, C. Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educ. Technol. Soc.* 2022, 25, 28–47.
50. Roll, I.; Alevan, V.; McLaren, B.M.; Koedinger, K.R. Improving Students' Help-Seeking Skills Using Metacognitive Feedback in an Intelligent Tutoring System. *Learn. Instr.* 2011, 21, 267–280.

51. Christodoulou, A.; Angeli, C. Using a Technology-Pedagogical Content Knowledge (TPACK)-Based Adaptive Learning System to Scaffold Student Learning. *Br. J. Educ. Technol.* 2022, 53, 567–588.
52. Parasuraman, R.; Manzey, D.H. Complacency and Bias in Human Use of Automation: An Attentional Integration. *Hum. Factors* 2010, 52, 381–410.
53. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arber, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* 2021, arXiv:2108.07258.
54. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; SAGE Publications: Thousand Oaks, CA, USA, 2018.
55. Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* 2007, 39, 175–191.
56. Braun, V.; Clarke, V. Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* 2006, 3, 77–101.
57. Ferguson, L.E.; Van den Broek, P.; Van Oostendorp, H. Adaptive Learning Environments and AI-Guided Instruction for Optimal Zone of Proximal Development Maintenance. *Comput. Educ.* 2022, 178, 104400.
58. Bjork, R.A.; Bjork, E.L. Desirable Difficulties in Theory and Practice. *J. Appl. Res. Mem. Cogn.* 2020, 9, 475–479.
59. Feng, S. Personalized AI Feedback in Programming Education: Effects on Cognitive Load and Learning Outcomes. *Comput. Educ.* 2025, 210, 105012.
60. Gkintoni, E.; Antonopoulou, H.; Sortwell, A.; Halkiopoulou, C. AI-Supported Cognitive Load Monitoring: A Systematic Review of 103 Studies. *Educ. Inf. Technol.* 2025, 30, 2345–2370.
61. Kruger, J.; Dunning, D. Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *J. Pers. Soc. Psychol.* 1999, 77, 1121–1134.
62. Paulhus, D.L. Measurement and Control of Response Bias. In *Measures of Personality and Social Psychological Attitudes*; Academic Press: San Diego, CA, USA, 1991; pp. 17–59.
63. Hofstede, G. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, 2nd ed.; SAGE Publications: Thousand Oaks, CA, USA, 2001.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.