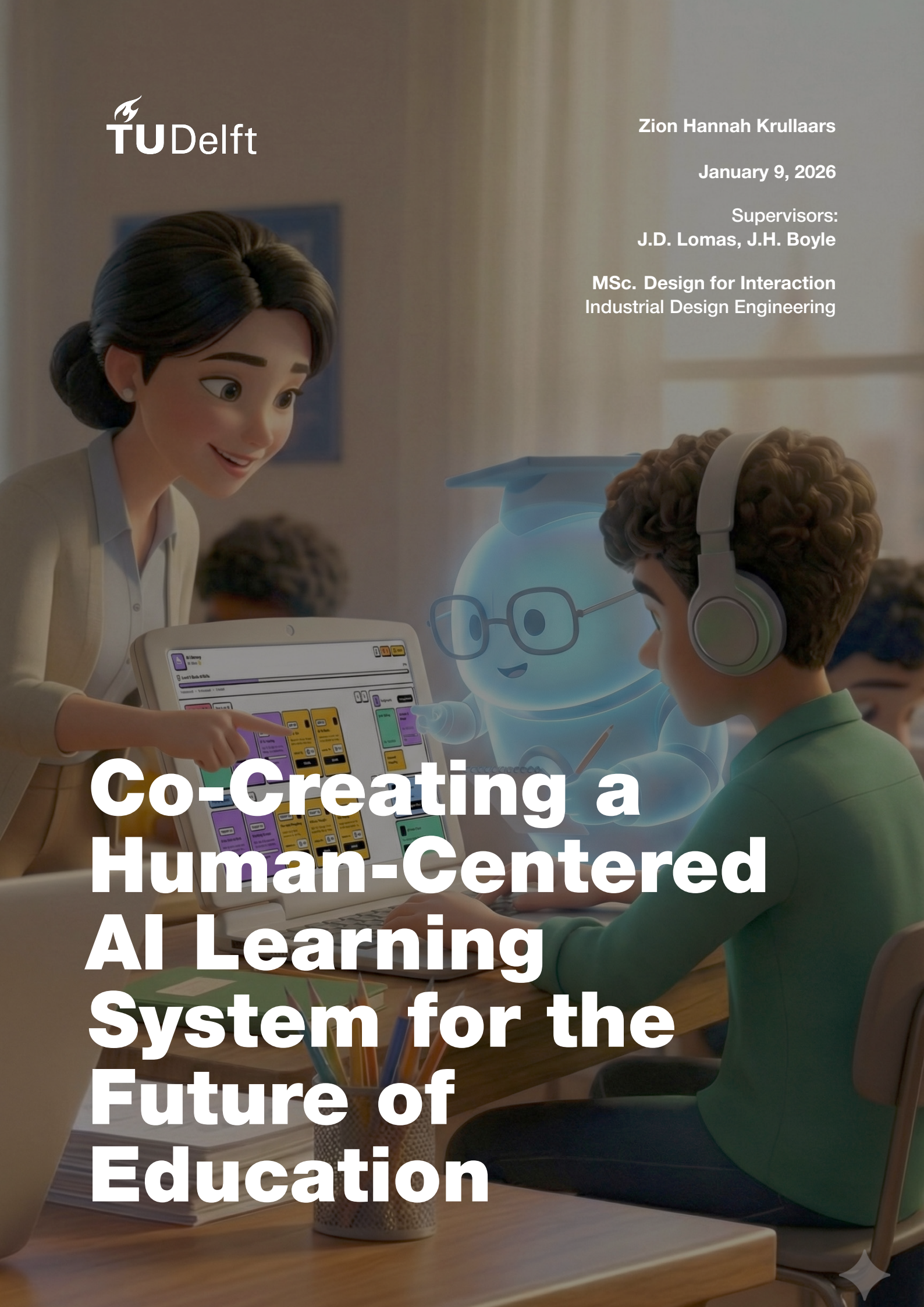


TU Delft

Zion Hannah Krullaars

January 9, 2026

Supervisors:
J.D. Lomas, J.H. Boyle

MSc. Design for Interaction
Industrial Design Engineering

Co-Creating a Human-Centered AI Learning System for the Future of Education

Co-Creating a Human-Centered AI Learning System for the Future of Education

Master Thesis
MSc. Design for Interaction
Industrial Design Engineering

Zion Hannah Krullaars

Internal Supervisors:
Prof. Dr. J.D. Lomas
Dr. J.H. Boyle

January 9, 2026

A.I. will force us humans to double down on those talents and skills that only humans possess. The most important thing about A.I. may be that it shows us what it can't do, and so reveals who we are and what we have to offer. **- New York Times**

Preface

This thesis is the final piece of a journey that started a very long time ago. It began during my first bachelor's project, where I explored the idea of "differentiation" through a project called "QuestCube," giving students the space to learn at their own speed. Since then, this theme has followed me like a common thread through several projects I've done.

While university sometimes feels like a bubble, my reality check often came from visiting the school where my mom works. Watching her teach and talking to her students reminded me that the "real world" does not always work the way I experience it. Not everyone uses AI or new tools the same way we do at university. This thesis was born out of a desire to bridge that gap and build something that truly works for teachers and students alike.

The path to finishing this was not always smooth. IDE is a "love-hate" relationship for me. I love the creativity, but I often resented it when it came to user testing. I used to dread it, only to find out it was never as bad as I feared (and secretly enjoyed it even). There were moments of total frustration, too. I will never forget spending an entire day having a breakdown over broken code, only to realize the problem was not my logic—it was just that my VPN was still on because I had been shopping for pizza stones earlier that day.

I hope this work serves as a foundation for introducing new technology into schools. Some of the findings here might seem "obvious," but I have learned that the most obvious things are often the ones we overlook.

Finally, I have to admit something to my younger self. For a long time, I took pride in saying, "I can do this on my own." But looking back, I know that is not true. I am fortunate to have had my parents, friends, and boyfriend by my side. On the days when even answering a WhatsApp message felt like climbing a mountain, their support (and sitting next to me until I had sent it) kept me going. This thesis may have my name on it, but I could not have finished it without them.

AI Statement

In the spirit of the very technology I explore in this work, I have used Artificial Intelligence (AI) extensively throughout the creation of this thesis. I used these tools as a "thought partner" to help brainstorm ideas, make mock-ups, and refine my writing.

I believe that using AI is a valuable skill, and I would highly recommend it to others as a way to enhance their work. However, I want to be clear: while AI helped me throughout, the heart of this project is mine. I have carefully checked, edited, and verified every part of this document. I take full responsibility for the final result and stand behind every word written here.

Executive Summary

As Artificial Intelligence becomes increasingly common in schools, many teachers worry about losing control over their classrooms[1–3], while students often use these tools as shortcuts rather than for genuine learning. This research addresses these challenges by aiming to co-create a human-centered learning system. Instead of viewing AI as a replacement for educators, the new structure is seen as a three-way collaboration between the Teacher, the Student, and the AI. The following pages summarize this approach, the development of the "Flight Simulator" teaching model (see Section 4.5), and the final design of a web platform that empowers students to learn safely while keeping teachers firmly in charge.

The Challenge and Vision

Today's education system faces a "one-size-fits-all" problem [4, 5]. Teachers want to help every student individually, but with 30 students in a class, it is physically impossible to customize lessons for everyone [6, 7]. Artificial Intelligence (AI) seems like the answer because it can personalize content [4, 8]. However, the current way AI is used in schools is flawed [9]. It is mostly "dyadic" (two-way): The Student talks directly to AI. The Teacher is left out of the loop. This creates a "black box". Teachers don't know if students are learning or just cheating (see Section 3.4). Students often use AI to get quick answers rather than to think critically (see Section 3.5). This leads to fear among teachers and lazy learning habits among students.

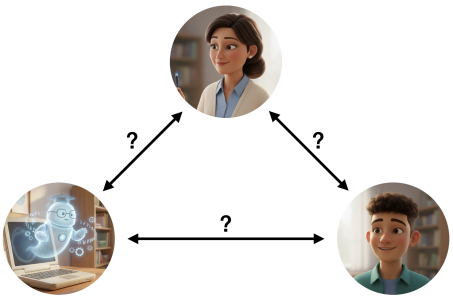


Figure 1: The three-way collaboration.

The Solution: Co-designed AI-integrated Learning System This thesis studies AI in the classroom with a focus on people. Here, the teacher is the leader, not a bystander. The student pilots their learning journey with AI as a helper, while the teacher ensures they succeed.

The Core Metaphor: The Flight Simulator To solve the conflict between "student freedom" and "teacher control" (see Section 1.2), this project uses the metaphor of a Flight Simulator: The Student is in the Cockpit. They are in the pilot's seat. They have a safe space to practice, make mistakes, and crash without real-world consequences. The Teacher is in the Control Tower. They are not sitting on the student's lap. They set the destination (learning goals) and watch the radar (data). They only jump in on the radio if the student is going off course. Lastly, the AI is the Co-Pilot. It sits next to the student, offering hints, warnings, and guidance, but it never flies the plane for them.

Key Takeaway AI should not replace the teacher. Instead, AI should handle the heavy lifting of logistics and basic explanations, freeing the teacher to focus on coaching, motivation, and human connection.

The Research Journey

The project did not just guess what users wanted; it went through three cycles of research with a "Think Tank" of real teachers and students (Section 3.1).

Cycle 1: Exploration (What is wrong now?) I interviewed teachers and watched students use current AI tools (see Sections 3.4 and 3.5). The teachers are afraid of losing control. They don't want to be "police officers" checking for cheating all day. They want to be coaches. Without guidance, students treat AI like a vending machine; they put in a question and want a direct answer. They trust the AI too much, even when it makes mistakes (hallucinations). Current AI tools are too polite. They give long lectures and don't challenge the students to think. Cycle 1 ended with a long list of requirements for the design of the future learning system.

Teacher-AI Interaction Requirements

- **T-AI 1:** **Content Control:** This is the most critical element. The teacher must have authority over the curriculum. The system should allow them to set learning goals and input their own materials. This ensures the AI's guidance aligns with the teacher's plan.
- **T-AI 2:** **Oversight without Micromanagement:** Teachers need to see how students are doing without becoming surveillance officers. A dashboard should offer high-level insights, highlighting struggling students so the teacher can intervene personally without reading every chat log.
- **T-AI 3:** **Process over Answers:** The AI must value the learning process. It should guide students toward an answer through inquiry rather than simply solving the problem.
- **T-AI 4:** **A Differentiation Engine:** To reduce workload, the AI should automatically adapt assignments and questions to different student levels, freeing the teacher to focus on high-value coaching.

Student-AI Interaction Requirements

- **S-AI 1:** **Balanced Conversation:** The AI should listen more than it talks, asking open-ended questions and keeping explanations concise.
- **S-AI 2:** **Guidance on "How to Use AI":** The interface should implicitly teach students how to prompt effectively and critically evaluate AI output.
- **S-AI 3:** **A Safe Environment:** The AI must provide a judgment-free space with positive reinforcement to build confidence.
- **S-AI 4:** **User-Driven Exploration:** The AI should allow spontaneous exploration without rigidly enforcing the initial lesson plan.
- **S-AI 5:** **Verifiable Accuracy:** The AI must rely on accurate, teacher-aligned sources to prevent the spread of incorrect information.

Foundational Requirements
User Experience and Interface (front-end):

- **UX 1:** **Intuitive Design:** Clear navigation and calls to action, usable without training.
- **UX 2:** **Visual Appeal:** A clean, professional interface that fosters a positive environment.
- **UX 3:** **Clear Progress:** Transparent purpose and visible progress indicators.
- **UX 4:** **Responsiveness:** Support across devices with on-demand assistance.

Core AI and Technical Capabilities (back-end):

- **TC 1:** **Pedagogical Model:** A dedicated teaching model supporting scaffolding and Socratic questioning.
- **TC 2:** **Robust NLP:** Accurate understanding of student input and teacher-provided materials.
- **TC 3:** **Learning Loop:** Continuous improvement through teacher and student feedback.
- **TC 4:** **Seamless Integration:** Tight backend integration without disrupting user experience.

Cycle 2: Design & Testing (Finding the Balance) Prototypes were built to test different levels of control. One prototype let teachers see every single word a student typed in real-time (see Section 5.1). Teachers hated this. It felt like spying and ruined the trust in the classroom. The next prototypes were based on "Structured Autonomy". The result was to give students a private safe space to talk to the AI. The teacher sees data (progress, mood, struggles), but not the private chat logs unless necessary. I asked students to try, and "break" the AI (e.g., asking it to plan a bank robbery or giving away the answers). The

AI was programmed to be resilient. When a student tries to trick it, the AI does not just say no. It pivots the conversation back to the lesson, turning the trick into a learning moment.

Key Takeaway Teachers do not want total surveillance; they want strategic oversight. Students need a private space to ask “dumb questions” without fear of judgment. The system must balance these two needs.

The Final Design

The result of the research is a web-based platform, called Cubo, with two distinct views: The Student Cockpit and The Teacher Control Tower.

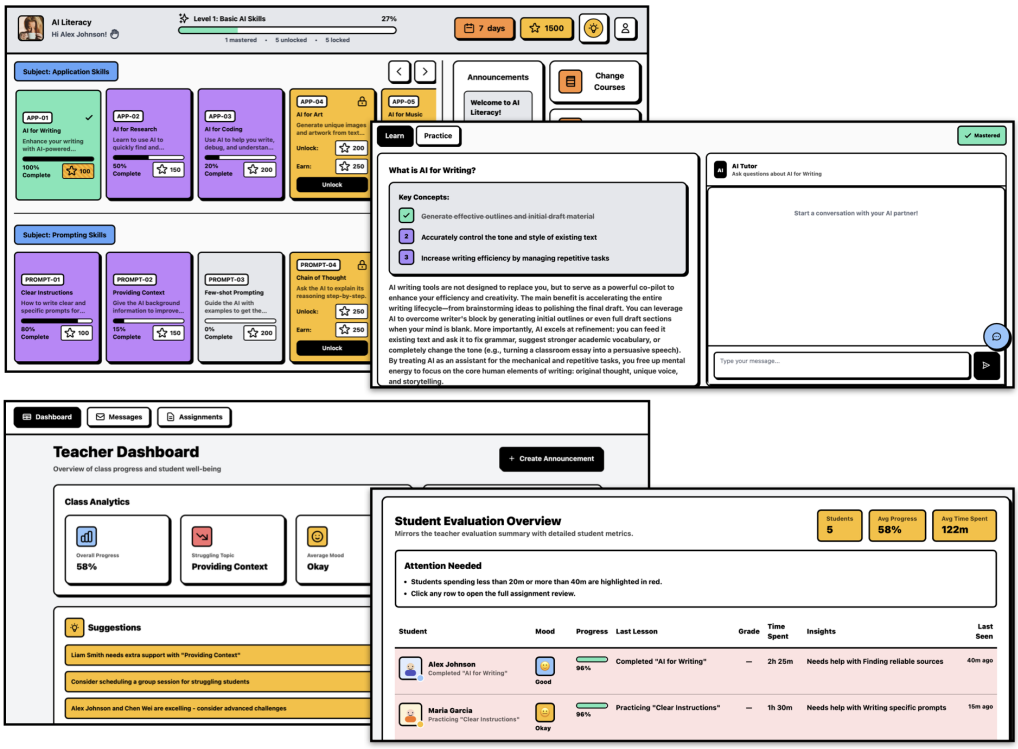


Figure 2: Screenshots of the student (top) and teacher (bottom) interfaces.

The Student Dashboard (The Cockpit) This is a game-like, personal interface designed to build “AI Literacy” skills. It is not an empty text box like ChatGPT. It has a progress bar and specific skill modules to break down the topic. To finish a lesson, the student can’t just read or simply ask for answers. They have to do things, like asking the AI for three different perspectives or fact-checking an AI claim. The AI is programmed to be a partner. It asks questions back (Socratic method [10]) rather than giving answers. The student enters their

interests (e.g., soccer, music). The AI uses these interests to explain difficult concepts using metaphors the student understands.

The Teacher Dashboard (The Control Tower) This dashboard let teachers experience the positive effects of AI without adding work. A quick view of who is falling behind and what topics the whole class is struggling with. The system alerts the teacher if a student is distracted, stuck, or moving too fast. This tells the teacher exactly who needs a human check-in. The teacher uploads their own lesson plans. The AI is programmed to follow this material, ensuring students reach the learning goals with the right content and jargon.

The Physical Box (Onboarding) Because new technology can be overwhelming, the project includes a physical "Starter Kit." It contains posters ("AI as Your Superpower"), a setup guide, and information about onboarding the students. This physical step helps ground the digital tool in the real world and gives the teacher ownership over the rollout.

Figure 3: A representation of the onboarding box.

Key Takeaway The design hides the complex technology behind a simple, friendly interface. It turns AI from a "cheat tool" into a skill-building tool.

Evaluation and Conclusion

To evaluate the system, I conducted a study with 16 students over 10 days. In this final experiment, one group used standard AI (like ChatGPT) and the other used the newly co-designed system “Cubo”. The results were clear:

Results Using the right tool made a big difference in student confidence. Students using standard AI actually lost confidence; they felt overwhelmed and unsure if the AI was lying (Figure 82, Condition A). In contrast, students using the new system gained significant confidence because it guided them step-by-step (Figure 82, Condition B).

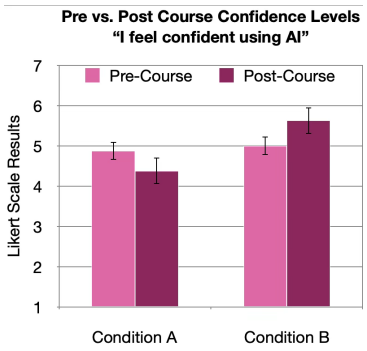


Figure 4: Pre vs. Post Course Confidence Levels.

The difference in confidence gain between the two groups was statistically significant. Students using Cubo learned to write better prompts, understanding the need to give the AI a role, context, and constraints. In terms of safety, 100% of students using the new system successfully caught the AI making up fake facts. In the standard group, only 75% of students managed to spot this misinformation.

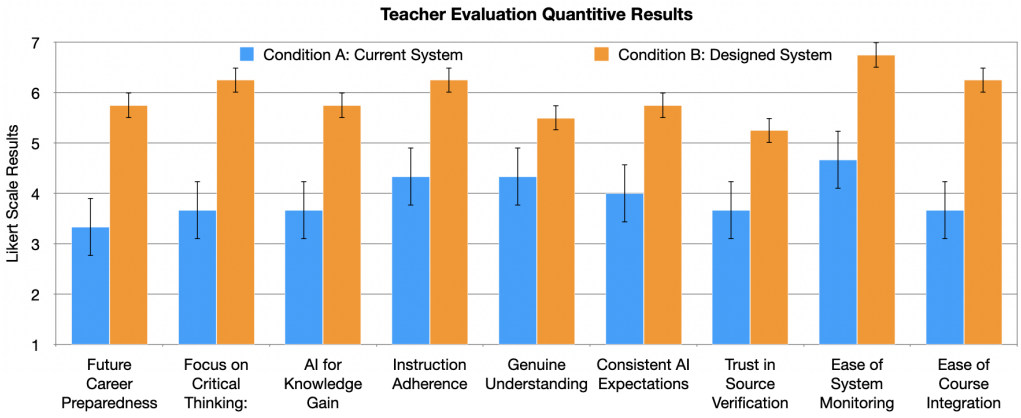


Figure 5: Teacher Evaluation: Condition B (Cubo) vs. Condition A over several themes

Teachers also rated the Triadic Tutor much higher for “Ease of Monitoring” and “Future Career Preparedness.” They felt less like police and more like coaches. The numerical data clearly show a preference for the new tool.

Conclusion The *Cubo* system proves we don’t have to choose between personalized learning and human teaching. By putting the teacher in the “Control Tower” and the student in the “Cockpit,” we can use AI to do the hard work of tailoring lessons to each student. This allows teachers to stop worrying about managing class logistics and focus on what machines can never do: empathy, mentorship, and shaping the personalities of their students.

Recommendations for future implementation (1) Before using AI for math or history, students must take a “driver’s ed” course on how to use AI safely. (2) Keep student chats private to encourage honest mistakes and learning. (3) Always ensure the teacher has the final say on the curriculum and content.

Final Verdict Cubo transforms AI from a threat into a partner, creating a future where technology makes education more human, not less.

Contents

01
Introduction

This section addresses the limitations of traditional ``one-size-fits-all'' education by exploring how Artificial Intelligence (AI) can enable personalized learning [4, 9, 11]. An informal survey confirmed that students feel a lack of control over their learning and find personalization options minimal, while teachers struggle to cater to diverse individual needs at scale [7, 12]. This creates a clear need for a new approach. Existing research on AI in education often focuses on one-to-one interactions between a student and a system [13], neglecting the complex classroom environment [14]. This project aims to find a co-designed solution to find and elaborate a new three-way partnership where a teacher, a student, and an AI work together collaboratively. The goal is to design a system that augments the teacher's capabilities, fostering a more adaptive and effective learning experience for students.

1 Introduction

During a peer review, a fellow student once accused me that my text “had to be fake” because I used an em-dash… The supposed calling card of AI-generated writing. The irony was that I had written every word myself. That moment stuck with me — it made me rethink not just how I write, but how deeply AI has already shaped our instincts and suspicions. Since the hype exploded in 2022 [1, 15], we began to ask AI questions, generate images, plan our days with it, and even sometimes just talk to it. I, too, began exploring the growing wave of tools reshaping how we learn and create.

When you examine how AI is impacting education today, you see that it is a total game-changer. Not only because of suspiciousness, but AI is fundamentally transforming education because it allows for personalized learning [4, 5, 11, 16], we can move past old-school, one-size-fits-all teaching. Today, intelligent systems can meet each student exactly where they are in terms of study progress, matching their specific needs and how they learn best. This type of education is called; Eduation 4.0 (Figure 6) [5, 17]. It helps solve the problem where teachers try to teach differently for every student but end up worrying that the weakest students are lost and the strongest ones are overlooked.

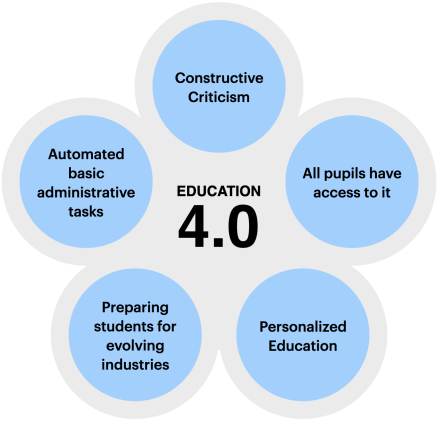


Figure 6: The five pillars of Education 4.0 by Forbes [17]. These are made with Industry 4.0 in mind. Industry 4.0 is ”the next phase in the digitization of the manufacturing sector, driven by disruptive trends including the rise of data and connectivity, analytics, human-machine interaction, and improvements in robotics.”, McKinsey & Company [18]

This new style of education can provide real-time feedback and dynamically adjust the course as you progress [8, 19]. But here, the most important part of this integration is to remember that AI is not replacing our teachers [20]. It is collaborating as a powerful partner that boosts what human teachers can do [21–23], making learning way more engaging and leading to better results for students, re-imagining education as it integrates into the system.

The current idea of AI in education is in the form of Intelligent Tutoring Systems (ITS) [8, 24]. Think of them as your own personal, digital mentor. These systems are prompted to help you reach your learning goals and adapt the lesson based on how you learn best [25], your personal speed, and all other specific help you need [11].

Before we can talk about where AI in education is headed next, I need to understand where it currently falls short, so that I know what actually needs improvement and where ITS can play a role. To uncover these first insights, the most natural place to start was simply to ask the people who work with these systems directly... students and teachers [3, 26, 27].

1.1 A First Impression of the Context

To understand the baseline of how personalized learning is experienced in practice, especially before the widespread integration of advanced AI tools, I conducted a brief, informal survey using Instagram Stories.

> Research Question - *To what extent do students perceive agency over their learning processes in traditional high school settings?*

The purpose was to invite participants to reflect on their experiences with personalization in high school, such as the freedom to choose topics, work methods, or pacing.

Informal Questionnaire

Method: I made seven Instagram story slides. The first one was to explain the research and inform about the use of the data. The other slides were questions that could be answered with sliders, open text blocks, and multiple-choice click boxes, see Appendix B.

Purpose: This informal poll invites participants to reflect on their experiences with personalization in high school.

Hypothesis: In the current educational system, people do not feel in charge of their own personalization, and the adaptation of the material to the student is minimal.

The survey was active for 24 hours and I got responses from 43 participants. The results highlight how education is experienced among the respondents. The insights from the respondents help paint a clearer picture of the state of personalized learning today and provide valuable context for the potential impact of AI-enhanced education.

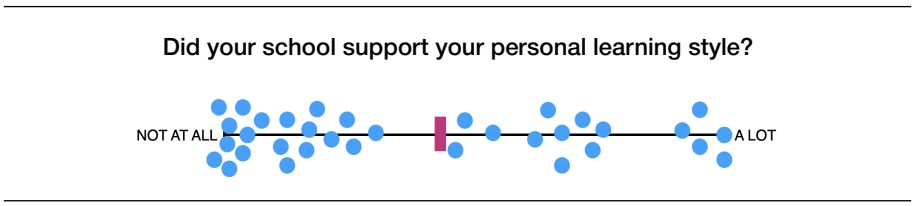


Figure 7: Results from the informal Instagram research. A blue dot indicates a reaction, the pink line the mean.

When asked if their school supported their personal learning style, the results were mixed. As shown in Figure 7, the responses were spread out across the scale, with the average (the pink line) landing near the middle. This suggests that the traditional school system is inconsistent: it works well for some students, but fails to support others. Similarly, when asked if they received enough guidance on difficult subjects (Figure 8), the results showed a wide range of experiences rather than a clear negative trend. This “middle-of-the-road” result highlights that while support exists, it is not tailored effectively to everyone’s needs.

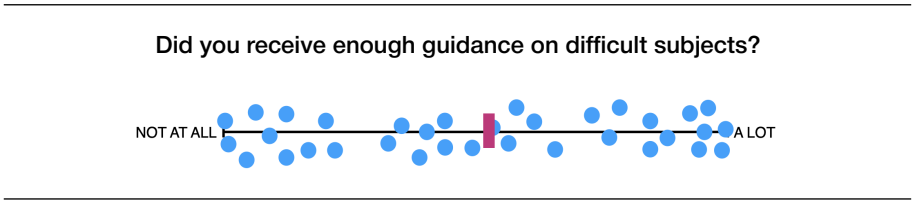


Figure 8: Results from the informal Instagram research. A blue dot indicates a reaction, the pink line the mean.

When asked who decides how and what they learned, the responses overwhelmingly pointed to the teacher and the school, with very few respondents feeling that they themselves were in control. This suggests a strong feeling of external control over the learning process (Figure 9). The available personalization options were scattered, with many respondents indicating "NONE" was available (Figure 10).

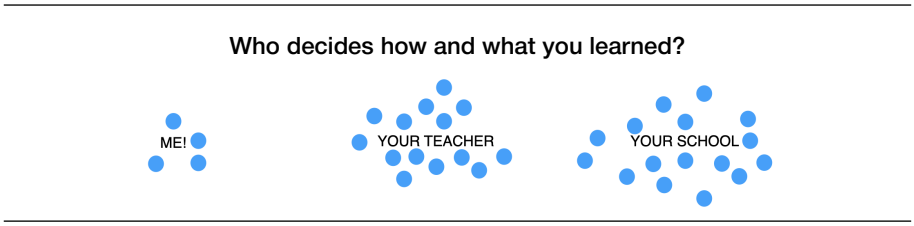


Figure 9: Results from the informal Instagram research. A blue dot indicates a reaction.

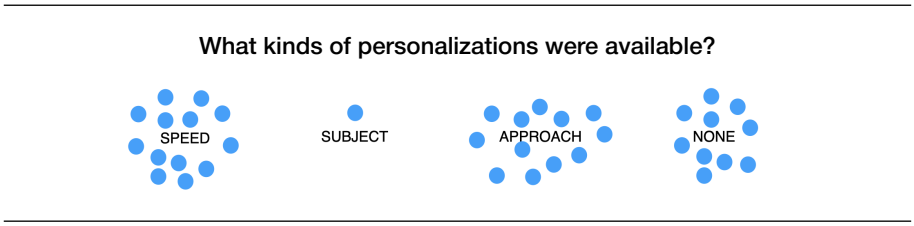


Figure 10: Results from the informal Instagram research. A blue dot indicates a reaction.

Qualitative feedback provided a deeper context. When asked how they could personalize their learning, answers were limited (Figure 11).

In what ways could you personalize your subjects?
"I could choose my subjects halfway through high school, but that's all."
"Mostly not, some teachers were creative and created groups"
"Not at all"
"The amount of homework that I wanted to do"

Figure 11: Qualitative results from the informal Instagram research.

When asked what they would like to change, both students and teachers expressed a desire for more dynamic and individualized methods (Figure 12).

What would you like to be different about the educational system?
"Some courses would have been more interesting if they were projects"
"Sometimes I was bored as the pace of the group was not aligned with my pace"
"As a teacher, I want students to work more on a project base"
"Sometimes when teaching I have difficulty focusing on all the students"

Figure 12: Qualitative results from the informal Instagram research.

The survey results, although informal, still strongly support my initial hypothesis. More importantly, they revealed several critical gaps that AI is perfectly suited to fill in education.

A Lack of Student Agency: The responses clearly indicate that students feel a lack of control over their own education. Learning is perceived as a process dictated by the institution ("your school") and the instructor ("your teacher"), rather than a collaborative or self-directed journey. This passive role can lead to disengagement, a point underscored by comments about boredom due to mismatched pacing.

"One-Size-Fits-None": The low ratings for support of personal learning styles and guidance on difficult subjects suggest that the traditional classroom model struggles to meet diverse individual needs. The open-ended desire for project-based learning and flexible pacing shows a demand for different approaches and speeds, two things that are difficult to manage in a standard classroom setting.

Teacher's Bottleneck: Crucially, the survey includes feedback from the teacher's perspective, highlighting the core challenge: it is incredibly difficult for one person to "focus on all the students" simultaneously. This is not a failure of teachers, but a limitation of the system. True personalization is not scalable through manual effort alone.

These insights helped me frame a really clear problem: students are seriously looking for more personalized learning, but teachers just do not have the capacity to deliver that to every single student. This is exactly the spot where AI-powered tools can step in. By automatically handling things like adapting content, giving feedback the second a student needs it, and managing all those individual learning paths, AI can give teachers the tools they need. This helps breaking past that old "one-teacher-to-many-students" bottleneck and actually create student-centered, adaptive learning environments that everyone in the survey was longing for.

1.2 Problem Context

The challenges the informal survey highlighted, things not being personalized enough and how one teacher needs to manage a diverse classroom, are exactly the issues that this thesis is focusing on.

The recent explosion of Large Language Models (LLMs) has accelerated this shift. They offer the potential to create truly dynamic, personalized learning paths for every student [4, 9, 25]. These technologies are the backbone of intelligent tutoring systems mentioned before, adapting to a student's pace and providing tailored resources in a revolutionary way that was not possible until before [8, 11].

As said earlier, there is a really important take: integrating AI into the classroom is not about replacing teachers. The emerging plan is all about human-AI collaboration, where the technology is a true partner to the educator [21, 28]. In this new model, the teacher's job evolves into that of an "orchestrator," someone who uses these AI tools to set up learning environments that are more effective and engaging [29]. This collaborative setup, involving the teacher, the AI, and the student, is central to actually getting the most out of this new educational technology [20].

Even with all these promises, we still have some significant hurdles to take. First, LLMs have technical limits. They can sometimes provide information that is biased or simply incorrect (also known as hallucination), so careful oversight is a must [22, 30–33]. Next there are pedagogical and ethical questions. A huge challenge is making sure that we use AI to help students develop critical thinking and not just let them cheat or have AI do all the assignments for them [34]. Plus, successful use also depends on overcoming the fact that many teachers might not feel confident using this new technology yet, meaning we need to include professional development [2]. Finally, there is the issue of fairness and access. The benefits of AI in education should reach as many students as possible, also those with special needs and limited access to resources [35].

These technical, teaching-related, and logistical challenges all underscore one thing: we cannot just expect schools to use AI tools. We need a system that thoughtfully integrates this technology into the existing classroom dynamic, making sure the teacher is involved every step of the way.

1.3 Research Gap: Co-Creating a New System

Research on AI in education has mainly focused on direct, one-on-one interactions between a person and a computer [9]. While this focus has produced useful results, it leaves an important gap in our knowledge: how to design systems that can handle the complex, connected reality of a real classroom [14].

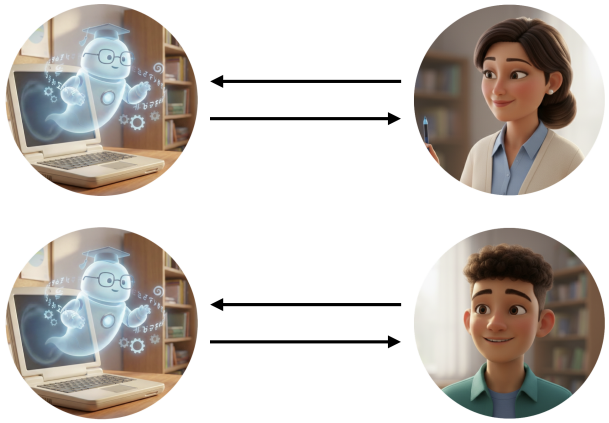


Figure 13: Current AI research often focuses on isolated, two-way interactions. Whether it is an AI answering a Teacher or quizzing a Student, these events happen separately. This setup often misses the bigger picture: the rich feedback and human connections that a fully integrated system requires.

This project addresses that gap by co-creating a human-centered learning system. I use the idea of a three-way collaboration as a tool to understand and design these interactions [20, 23]. This perspective helps me move beyond simple tools and build a partnership where the AI acts as a supportive participant in the learning process [20, 21].

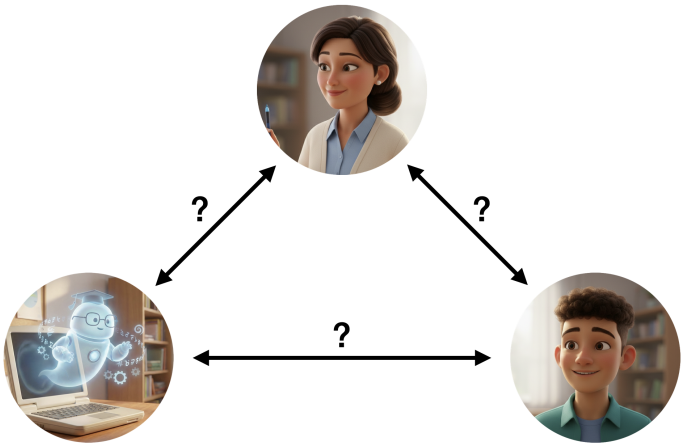


Figure 14: The proposed human-centered system, viewed through a triadic lens. Here, the AI Agent, Teacher, and Student work in a shared loop. I use this framework to ensure the AI supports the flow of information between humans rather than interrupting it.

1.4 Research Questions and Design Process

The primary objective of this project is to architect, design, and validate a Human-Centered AI learning system. Rather than merely observing the theoretical implications of AI in education, this research aims to a three-party collaboration into action to facilitate a symbiotic ecosystem involving the teacher, the student, and the AI. To guide this inquiry, the research is structured around three Main Research Questions (MRQs):

MRQ 1: *To what extent do current dyadic (one-to-one) LLM interactions satisfy the pedagogical requirements of personalized learning, and where do they fail to account for the holistic classroom context?* This question assesses the foundational efficacy of LLMs in current ITS. I will establish a baseline by critically evaluating the capacity and limitations of existing Generative AI (LLMs) within a simple, bilateral user-system dynamic. This involves quantifying the system's current ability to generate personalized content and collaboration based on initial prompts, while simultaneously analyzing stakeholder experiences to identify friction points in the current implementation.

MRQ 2: *What specific interaction modalities and systemic features are requisite to transition from the current diadic situation to a collaborative AI-integrated learning system?* I will set out the systemic requirements for a new co-created learning system that incorporates AI. Moving beyond the dyadic model requires understanding the current interactions and the needs, limitations, and preferences from the parties involved. This question focuses on identifying the critical architectural gaps and defining the essential feature set required for the AI to function not merely as a repository of information, but as an active co-pilot within the learning process.

MRQ 3: *How can a co-created, human-centered platform effectively orchestrate the feedback loop between Teacher, Student, and AI to ensure pedagogical control remains with the educator?* The ultimate goal is to synthesize the findings into a tangible tool that works for everyone. Based on the research, this means building a system where teachers can easily monitor progress and students feel safe and guided. By addressing these specific needs, we can ensure the tool is not just theoretically good, but something users will actually want to adopt.

To address these questions, I use a user-centered design approach with a strong focus on participatory co-creation. Rather than developing the system in isolation, it is designed together with the people who will use it in the classroom. Teachers and other stakeholders are actively involved throughout the process, and the system is repeatedly tested in realistic educational settings. This ensures that the resulting tool closely aligns with everyday teaching practices and constraints [36, 37].

The project follows a three-stage iterative research process, shown in Figure 15. The first stage explores the broader framework and stakeholder landscape, identifying key actors and mapping interaction flows to form an initial vision of the product–service system. The second stage builds on this by developing and refining prototypes, which are tested through

multiple iterations in real-world contexts to evaluate the feasibility of the triadic approach. The final stage focuses on evaluating and refining the newly designed system.

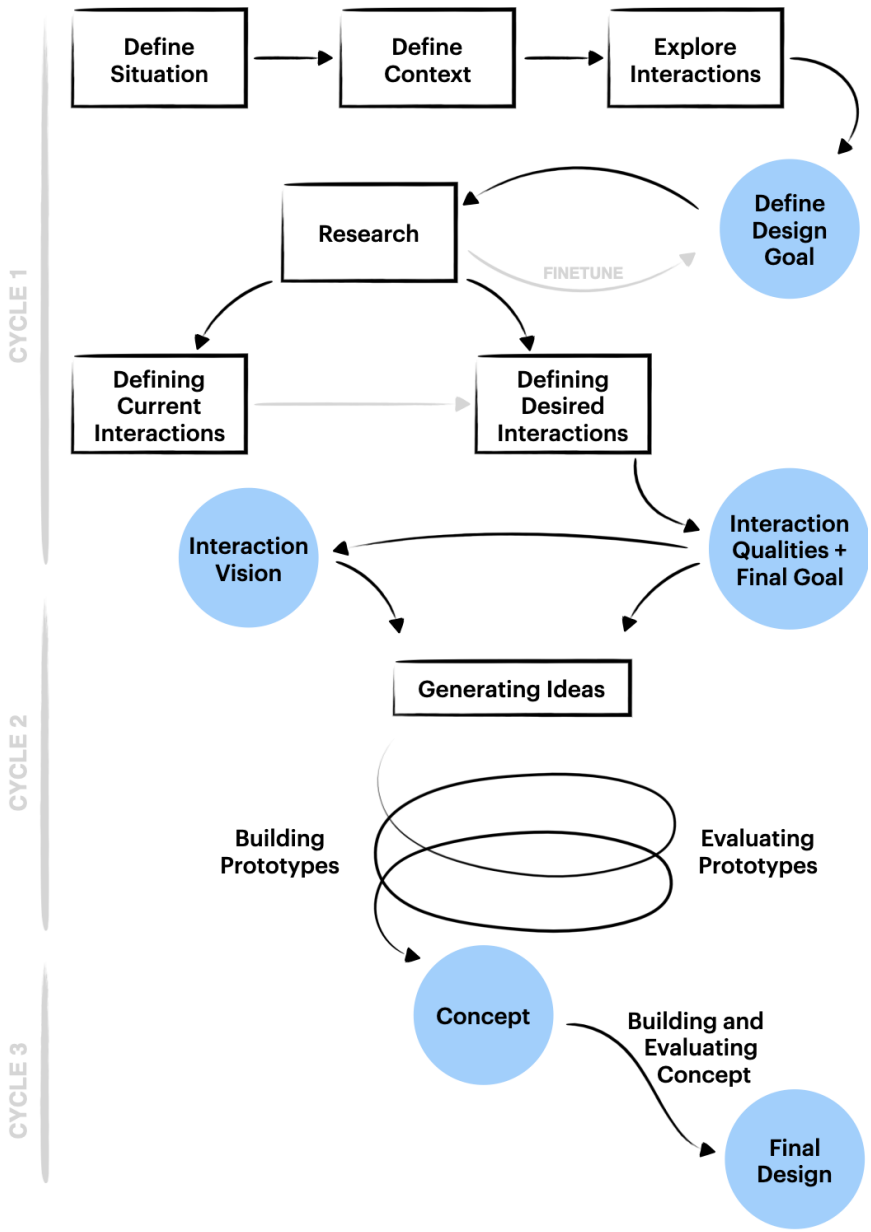
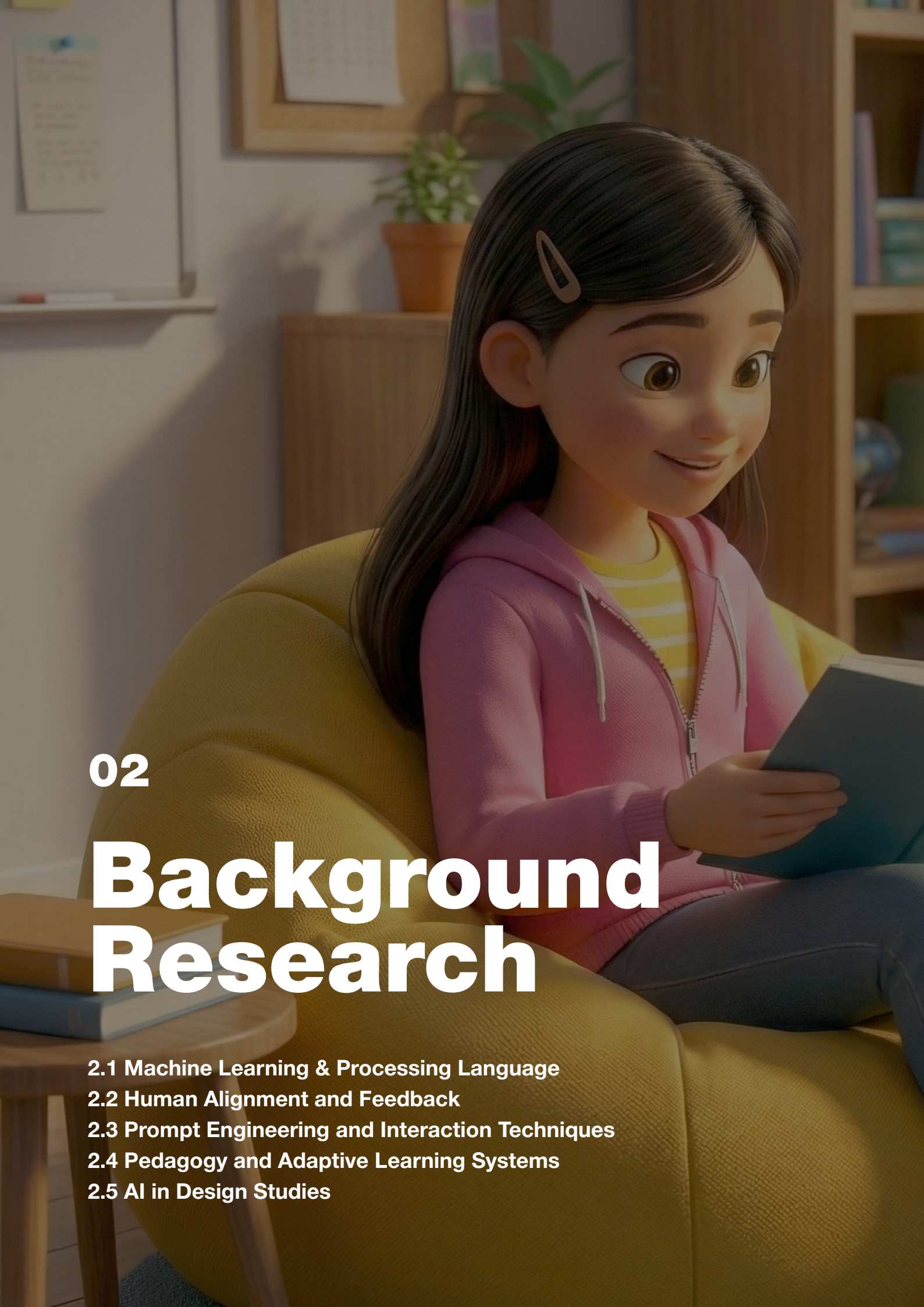


Figure 15: The comprehensive interaction design methodology guiding this thesis. The process comprises three distinct phases: Phase I explores current interaction dynamics; Phase II focuses on future-state modeling and prototyping of the system; and Phase III evaluates the efficacy of the final design.



02

Background Research

2.1 Machine Learning & Processing Language
2.2 Human Alignment and Feedback
2.3 Prompt Engineering and Interaction Techniques
2.4 Pedagogy and Adaptive Learning Systems
2.5 AI in Design Studies

The adoption of Artificial Intelligence, specifically Large Language Models (LLMs), is shifting education from a ``factory model'' to a personalized one, often called ``Education 4.0'' [5]. However, simply dropping AI into a classroom is not enough. To build a system where the teacher is in charge and the AI acts like a tutor, we first need to understand the mechanics of the technology. This chapter breaks down how these models ``think,'' how we can train them to be safe, and how we can design them to support human interaction [4, 9].

2 Background Research

2.1 Machine Learning & Processing Language

To understand how an intelligent tutoring system operates, we must first define the essential building blocks of Artificial Intelligence (AI) and Machine Learning (ML). These are the fundamental mechanisms that allow a system to process information, make decisions, and integrate with other software.

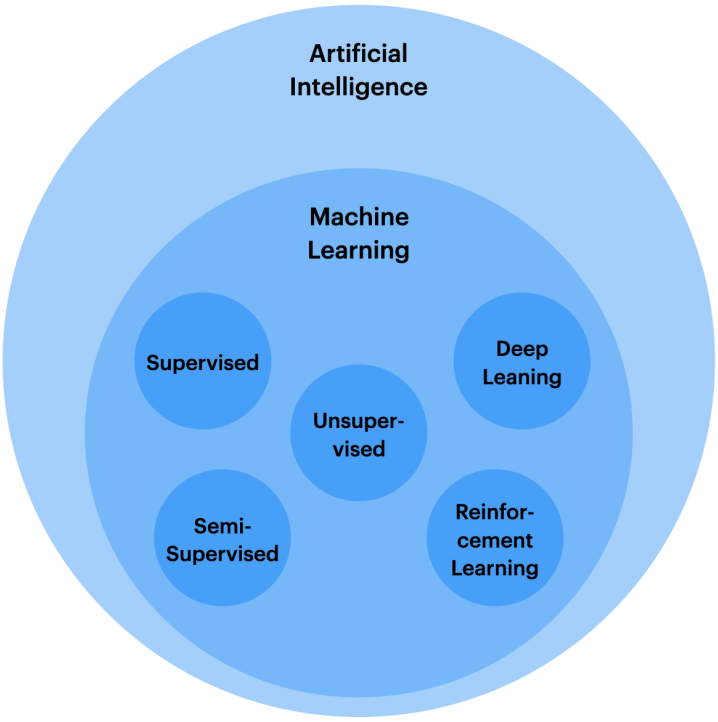


Figure 16: Artificial Intelligence is a program that can sense, reason, act, and adapt. Machine Learning is a subset of AI where algorithms improve as they are exposed to more data. This field further branches into supervised learning, unsupervised learning, deep learning, and reinforcement learning. [38, 39]

At the core of modern AI is the model [40]. Think of the model not as a static database, but as a program that has been trained on vast amounts of data. Its primary function is to make predictions or decisions based on what it has learned, typically executed through an algorithm, a precise set of rules or steps the machine follows to complete a specific task [39]. To make this intelligence accessible to a learning platform (like the dashboard designed in this thesis), we rely on an API (Application Programming Interface). The API acts as the bridge, providing the definitions and protocols that allow different software components to communicate seamlessly [41].

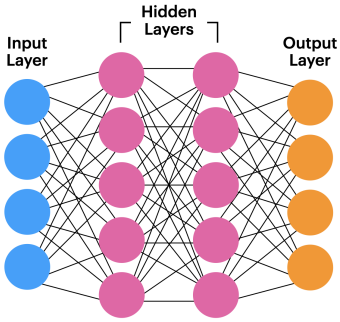


Figure 17: Layers in a neural network architecture: (1) The Input Layer receives the initial data, where each node corresponds to a specific feature. (2 & 3) The Hidden Layers perform the computational "heavy lifting," transforming inputs through multiple stages. (4) The Output Layer produces the final result, which depends on the specific task (e.g., predicting a student's grade). [42]

Delving deeper into the architecture, the most influential structure in modern AI is the neural network. Loosely modeled after the human brain, it is composed of interconnected layers of nodes where the strength of the connections is controlled by a "weight" [43]. An AI model is essentially only as good as its training; the rigorous process of optimizing performance by iteratively adjusting these internal weights based on data [43].

The complexity of these networks depends on whether the mathematical functions within the layers are linear or non-linear [43]; when a network becomes sufficiently massive, it is often referred to as a "black box" model because its internal decision-making process becomes difficult to interpret [44]. To judge the success of a trained model, we use a performance metric like accuracy, but crucially, we must test for generalization using instances the model has never encountered [45]. Here, data quality is paramount; an imbalanced dataset, where classes are unequally represented, can lead to significant bias [46].

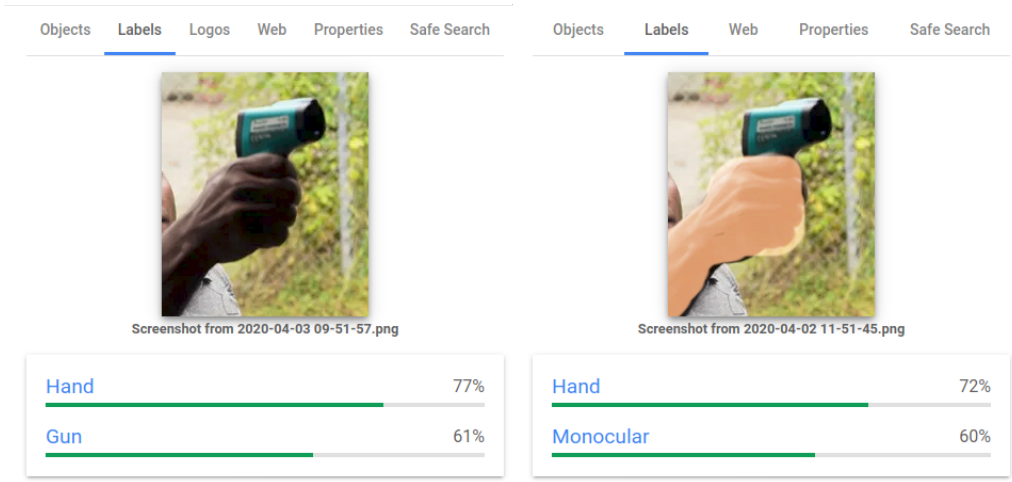


Figure 18: The Consequences of Bias: This visualization [47] shows a failure in model generalization. An infrared thermometer is misclassified based on a protected characteristic (race) due to unrepresentative training data. Ensuring data equity is essential to prevent such harms.

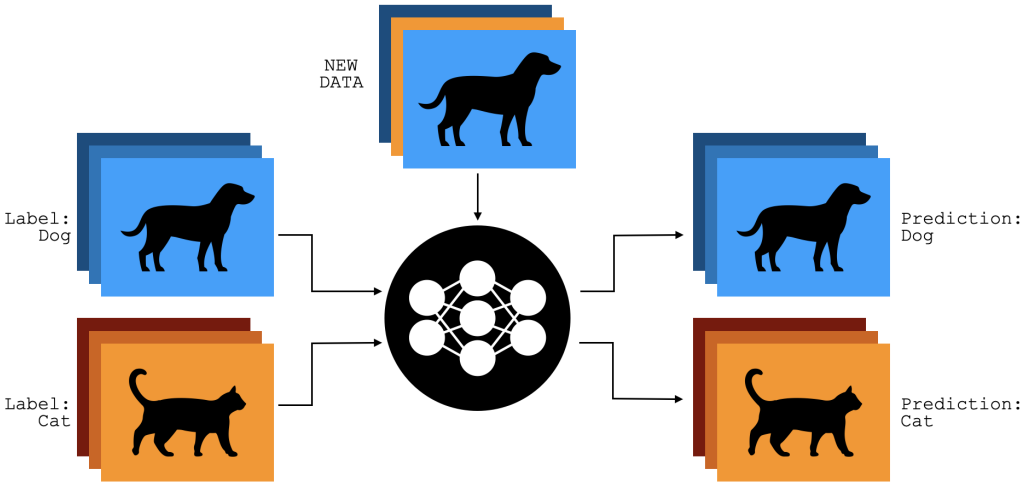
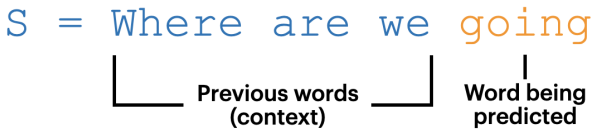


Figure 19: Classification Training: A neural model is trained using labeled data. This data will help the model understand the two classes. However, if one class (e.g., 'Dog') is over-represented, the dataset is imbalanced and the model can become biased.

The primary goal of many such models is classification, where the system assigns an input to a specific category (e.g., flagging a student's response as 'correct' or 'incorrect') [43]. Once deployed in a real-time classroom setting, latency becomes a critical metric; this represents the delay between a user's request and the AI's response [45]. Models can be deployed as offline/static models (fixed after training) or online/dynamic models, which continuously update their parameters as new data arrives [45].

While neural networks provide the structure, the ability to converse with students and teachers relies on a specialized field: Natural Language Processing (NLP), revolutionary architectures that allow machines to interpret and generate human language. This field is defined by Language Models (LMs), which are probabilistic engines designed to predict and generate text [48].



$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

Figure 20: LMs use predictive power to generate realistic-sounding human language. However, it is crucial to note that the model relies on identifying statistical patterns, not true human-like understanding of the semantic content. [49]

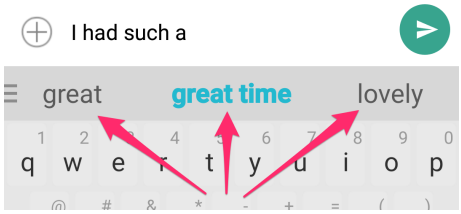


Figure 21: A simple example of word prediction. This probabilistic mechanism is what allows the AI to complete sentences in a coherent way.

The most powerful iterations, Large Language Models (LLMs), contain billions of parameters and drive Generative AI, the kind of AI that you see all around you nowadays that is capable of creating new content [50]. The pivotal breakthrough enabling this progress is the Transformer architecture, which underpins models like GPT and Gemini [51, 52].

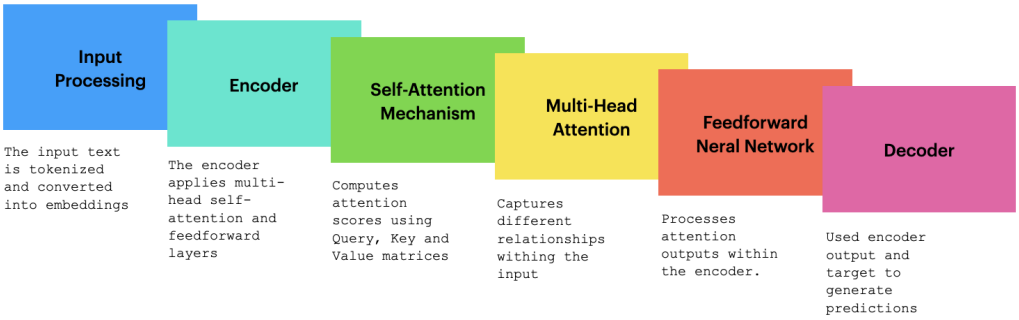


Figure 22: The Transformer model consists of an encoder (for understanding) and a decoder (for generating). It begins by tokenizing the text and assigning order information. The encoder uses self-attention to weigh the importance of each word relative to the others, understanding the full context. The decoder then generates new text using masked attention, predicting the next word based solely on previously processed context to improve accuracy. [51]

However, machines do not read words; they process numbers. The models use tokenization to break text into manageable units, or tokens. A model's short-term memory is defined by its context window, the maximum number of tokens it can process in a single interaction [52].

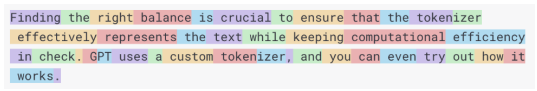
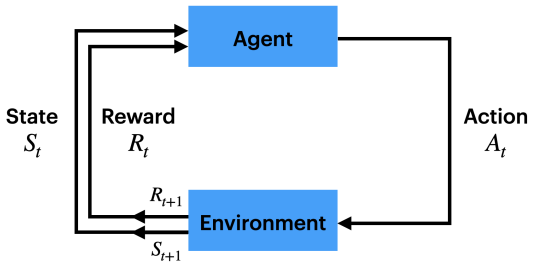


Figure 23: Tokens are not merely words or letters; they are semantic units. A tokenizer breaks a sentence into parts that help the model infer meaning.

2.2 Human Alignment and Feedback

An AI model can be trained on a lot of data, but it can also learn from the environment via rewards. When this is the case, it is called reinforcement learning (RL). This subpart of ML provides the framework for decision-making in complex environments. RL is about agents learning through trial and error. An agent performs actions in an environment to maximize a cumulative reward, a process mathematically modeled as a Markov Decision Process (MDP) [53]. The agent's strategy is defined by its policy, which maps the current situation (state) to the best action.

Figure 24: A visualization of a MDP. An Agent (e.g., an AI playing a game) takes an Action (A_t) within an Environment (e.g., the game world). The Environment then transitions to a new State (S_{t+1}) and provides a Reward (R_{t+1}) to the Agent. This continuous interaction of observing states, taking actions, and receiving rewards allows the Agent to learn an optimal policy to maximize cumulative rewards over time.



In an educational context, a raw model is insufficient; it must be safe, factual, and aligned with human values. This necessitates advanced alignment techniques that keep humans in the loop. A primary method for this is Reinforcement Learning from Human Feedback (RLHF), where the model is fine-tuned based on the rewards a human gives [54, 55]. The feedback can be dense (rewards given at every step) [56, 57], or sparse (rewards given only at the end of a sequence). This information is used to training a Reward Model (RM) to predict the score a human would give to a specific response [58, 59].

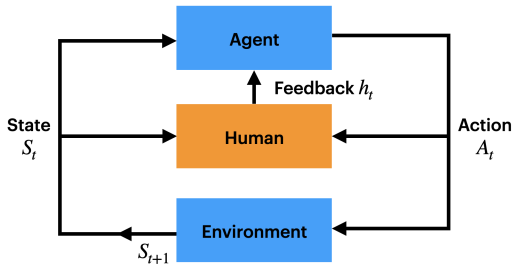


Figure 25: RLHF Integration: This graphic illustrates how human expertise is injected into AI training. The AI agent acts, and a human provides feedback. This feedback acts as a shortcut, explicitly telling the AI how to improve, leading to safer and more aligned systems.

Alignment also focuses on ensuring trustworthiness, a non-negotiable factor in an educational setting where teachers must rely on the system's accuracy. Consequently, factuality (the truthfulness of generated information) and groundedness (ensuring the output is supported by verifiable source material) become critical requirements for any AI tutor [22, 33].

Advanced algorithms like Proximal Policy Optimization (PPO) and Direct Preference Optimization (DPO) are often used to refine LLM models [60, 61]. In the context of this thesis, these algorithms can facilitate Multi-Agent Learning (MARL), enabling the system to learn from the interactions between multiple agents (e.g., a student and a teacher). This creates complex, supportive dynamics that evolve to become increasingly personalized over time [62, 63].

2.3 Prompt Engineering and Interaction Techniques

Once the model is trained and aligned, the challenge shifts to interaction. Prompt Engineering is the art of effectively communicating with LLMs to elicit specific, high-quality outputs suitable for a classroom environment.

The interaction begins with the prompt (the input) and ends with the response. Prompt engineering involves designing these inputs strategically [64]. A common technique is role prompting, which instructs the model to adopt a specific persona, such as "Act as a historian," to tailor the tone and complexity of the answer [65].

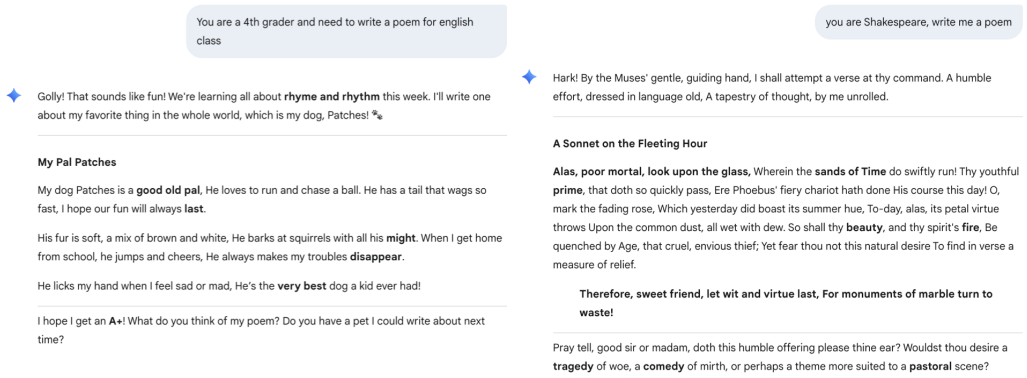


Figure 26: Versatility in Output: Whether generating a simple poem for a child or a complex one in the style of Shakespeare, the model adapts based on the prompt. Learning to prompt effectively is key to retrieving the desired output.

A powerful capability of LLMs is in-context Learning, where the model learns from examples provided directly within the prompt [48]. This spectrum includes Zero-Shot Learning (no examples), One-Shot Learning, and Few-Shot Learning (multiple examples). Additionally, Chain-of-Thought Prompting encourages the model to explain its reasoning steps before giving a final answer, which is particularly useful for tutoring math or logic [65]. Newer concepts like "vibe coding" refer to using AI to generate entire software components based on descriptive elaborate prompts [66].

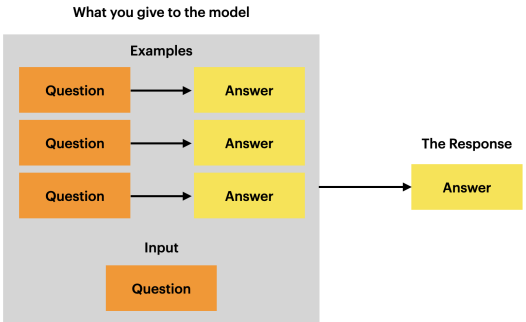


Figure 27: Few-Shot Learning: By providing Question-Answer pairs within the prompt, we give the model in-context guidance. This mini-training set helps the model understand the desired format and logic before it attempts to answer the user's question.

2.4 Pedagogy and Adaptive Learning Systems

The ultimate goal of the technologies described above is to serve a pedagogical purpose. In this thesis, I want to create an adaptive learning system that support students effectively. Optimal systems use AI to modify content, pace, and scaffolding based on individual performance [8, 67]. The objective is to keep the learner in the Zone of Proximal Development (ZPD); the "Sweet Spot" where a task is challenging enough to be engaging but achievable with support [67]. This aligns with experiential learning, which emphasizes learning through reflection on doing [68].

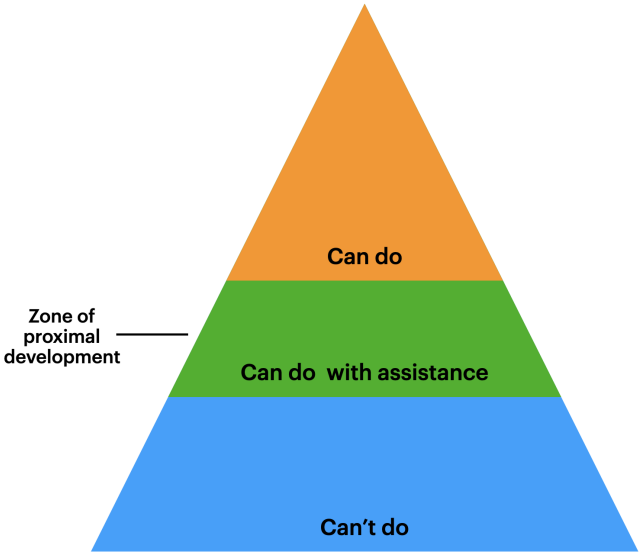


Figure 28: The Zone of Proximal Development (ZPD): Teaching below the ZPD leads to boredom, while teaching above it leads to frustration. The goal of AI scaffolding is to keep the student in the middle zone, providing just enough guidance to help them reach learning outcomes independently. [69]

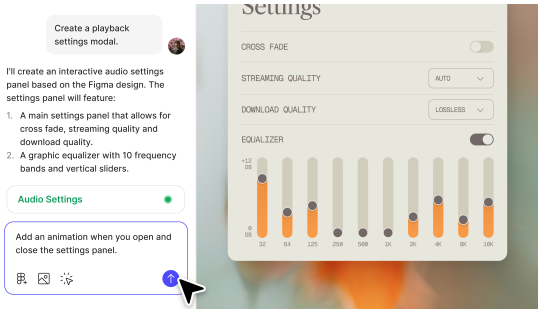
Designing these systems requires careful stakeholder alignment. We must balance the often-conflicting goals of different parties (teachers, students, administrators) [70, 71]. In technical terms, this mirrors "gradient conflicts" in model optimization, where improving one metric might degrade another. The challenge is to find an acceptable trade-off that maintains trust and effectiveness.

Figure 29: Conflict of Command: When an AI receives contradictory instructions from different stakeholders, it faces an alignment challenge. Clear guardrails are essential to resolve this friction.

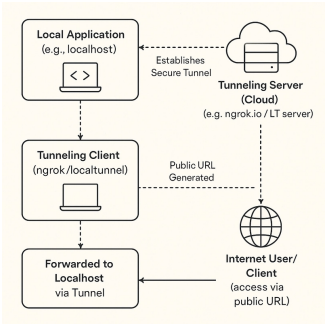
2.5 AI in Design Studies

The integration of AI into the design research process can significantly accelerate development [72]. In this thesis, AI tools are not only the subject of study but also the primary instruments used to build the system itself. This section outlines the technical stack and methods used to move from abstract pedagogical ideas to a functional prototype.

The process begins with generative ideation, using LLMs such as ChatGPT and Gemini to explore the classroom dynamics described in Section 3. Through targeted brainstorm prompting, I rapidly produced prototypes in which teachers, students, and AI might come into conflict, helping refine the "interaction vision" before any coding took place [73].



Figma Make interface: Generating the front-end visual structure based on simple text prompts.



Tunneling exposes a local development server to the internet via a secure bridge, allowing for the testing of AI agents without complex deployment setups [74].

Next, the project shifts from text to visuals using Figma Make. AI features enable the quick generation of high-fidelity mock-ups and wireframes from text descriptions [75]. This step is essential for shaping the dashboard that makes the triadic model tangible and ensuring the interface is usable for non-technical teachers [36].

Finally, the development phase uses AI-assisted coding to bridge design and implementation. The prototype is built in React [76], but instead of hand-coding every component, the project uses Cursor, an AI-first code editor that helps generate the complex logic behind the multi-agent system [77]. For local testing of AI agents and their communication with external webhooks, NGROK provides secure tunnels [78]. The full codebase is version-controlled through GitHub, supported by AI coding assistants [79].

03

Cycle 1: Exploring the Current Interactions

The previous chapters established the theoretical foundations for this project, defining the limitations of current educational AI and introducing the idea of a three-party collaboration. However, the ultimate goal is to co-create a tangible Human-Centered Learning System, not just a theoretical idea. To move this vision from concept to practice, I initiated the first cycle of my research: a broad, exploratory design journey. This cycle focused on understanding the current interactions between teachers, students, and technology to ground the new system in real experiences.

3 Cycle 1: Exploring the Current Interactions

Building on the technical basics we covered, this project focuses on co-creating a Human-Centered AI Learning System. I use the triadic model (Figure 14) as a design tool to put this system into practice. This approach moves past the old idea of AI just being a passive utility. Instead, it fosters a true partnership where both the human and the AI act as co-learners and co-teachers to reach shared educational goals [28]. This recognizes that good teaching involves complex interactions between several people and systems, all with different roles.

Learning is naturally social and collaborative. We build knowledge by interacting with others, and the quality of these interactions highly affects what we learn [80, 81]. By applying the triadic lens, the system includes AI as a real participant in this process of building knowledge, while still ensuring that human relationships remain the foundation of education.

3.1 The Approach

I used a human-centered, repeated design process, as shown in subsection 1.4, Figure 15. This cycle is not a straight line; it is a continuous loop of learning where what I find in one step helps shape the next one. Before I could write any code or design any part of the system, the first absolutely crucial step was to ground the project in the real experiences of the people who would actually use it. The success of this new system would not be about how smart the technology was; it would be about whether teachers and students would actually use and accept it in their classrooms.

My goal for this initial phase was clear: I needed to stop making assumptions and instead talk directly to teachers and students to understand their current relationship with technology, what they needed for teaching, what is frustrating them, and what they are hoping for. This first design step was all about listening, observing, and defining the problem. Not from my perspective, but from theirs.

That is why I set up a Think Tank, a group of 28 people that will be involved through various stages of the project. They will test prototypes, take part in co-design sessions, and share their experiences. This group consists of both teachers and students (all 18+). The participants were selected to be a diverse group with different backgrounds to gain insights from various angles. Figure 32 shows the demographics of the Think Tank.

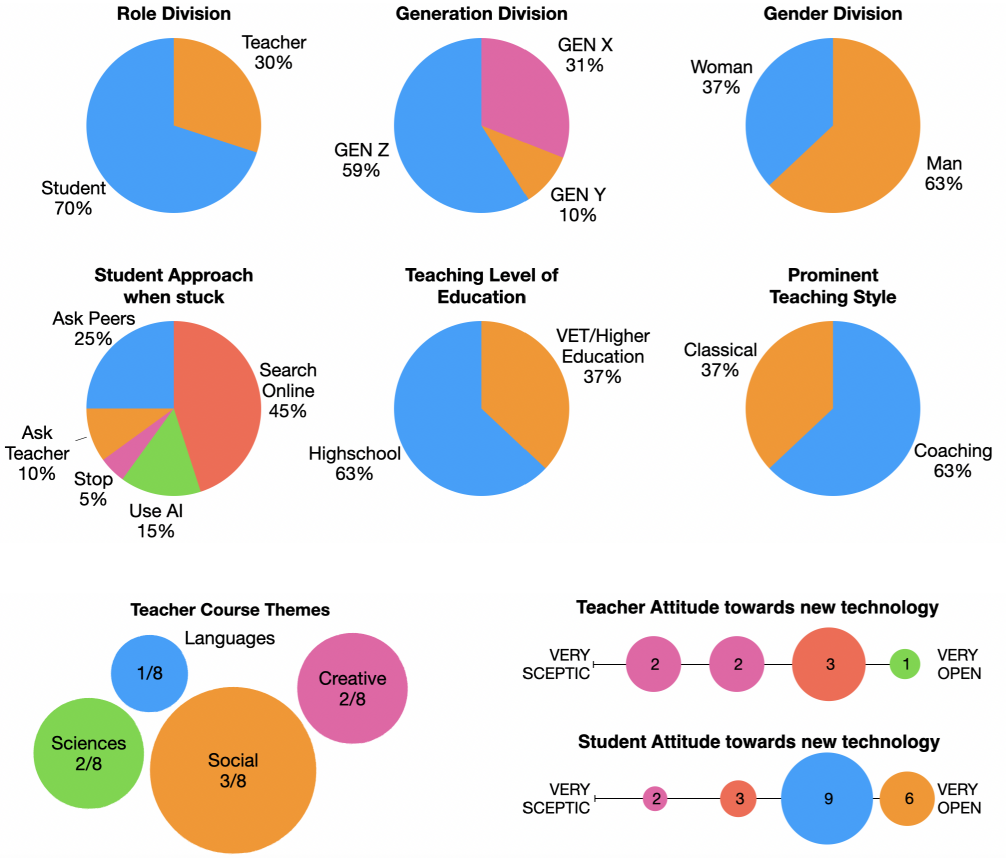


Figure 32: The demographic within the ThinkTank WhatsApp Groupchat. Each participant has filled in a consent form to participate in the multi-staged research.

3.2 The First Prototype Tests

To begin the exploration, I needed a shared object for discussion, something on paper that teachers could react to. This is why I created a simple online pixel game. This game included the first, and most obvious, mechanics to get a first tangible idea of the interactions.

> Research Question: *What features do teachers want in game-based learning tools?*

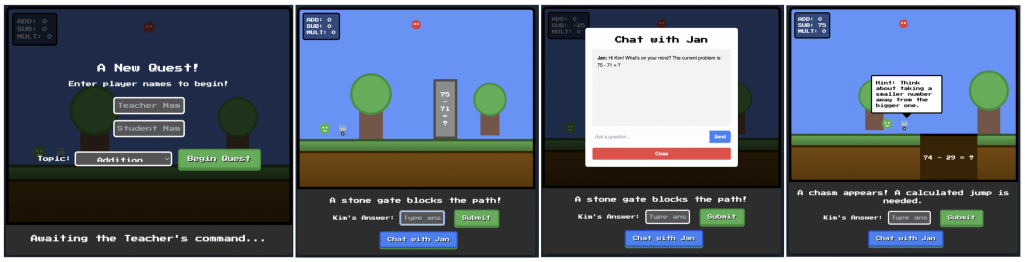


Figure 33: Screenshots of the first triadic prototype. Screenshot [1] initialization of the topic by the teacher, [2] shows the environment with the obstacles, [3] shows the interaction with the teacher, and [4] the prompted extra explanation when something is wrong

In the pixel game, a student would walk through an online world, solving problems along the way. They get tips from an AI tutor and can ask the teacher questions, but point will be reduced to stimulate independence.

Pixel Game Baseline Test

Method: A simple, functional web application was shared with the 2 teachers of the Think Tank, who were asked to explore it and provide their first reactions.

Purpose: To get a baseline understanding of the mechanics of the triadic system and what is important for teachers.

Hypothesis: The teachers will like the game environment and think that it will help students spend more time on their homework.

Testing this first tangible idea gave me a clearer view of the teachers' perspective. I initially thought that deducting points for help would encourage independence, but the teachers insisted that students must always feel free to ask questions. While they appreciated how the hints worked, they noted a missing feature: the ability for students to explain their reasoning. Finally, they cautioned that the design might be too playful. This raises an interesting challenge for me: how to make the game fun without losing the serious focus students need to learn.

3.3 The Baseline of Prompted LLMs

After testing the first game, which was mostly focused on the mechanics of a triadic collaboration, I had a lot of questions concerning the integration of prompted LLMs in educational settings. This led to the ideation and rapid development of a "Simple LLM Game," a minimal prototype designed to get a baseline reaction to the core idea of the triadic model with the integration of a prompted LLM.

> Research Questions: *How do students experience a prompted LLM interface? What level of control do teachers need over AI tutoring content?*

LLM Prototype Test

Method: A simple application was shared with the 28 members of the Think Tank, who were asked to use it and provide feedback through follow-up questionnaires and direct observation.

Purpose: To get a baseline understanding of how a simplified triadic interaction is perceived by all user types and to gather broad initial feedback on the core mechanics of a collaborative learning tool.

Hypothesis: Users will appreciate the core conversational and adaptive nature of the AI tutor but will identify significant gaps regarding teacher oversight and pedagogical depth.

The prototype was very straightforward: A teacher inputs a learning objective, and the AI engages the student in a conversational game to achieve it, some screenshots of such interaction can be found in Figure 34.

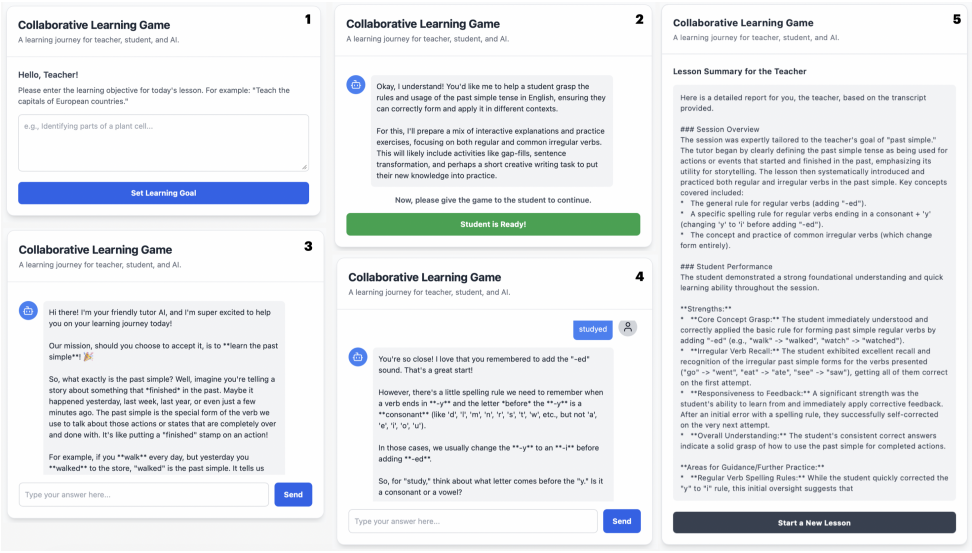


Figure 34: [1] The teacher interface to explain the assignment, [2] The confirmation by the chat bot, [3] The first interaction with the student, [4] The prompted positivity and extra explanation when something is wrong, and [5] The lesson summary as a reflection for the teacher.

This allowed me to directly test a simplified version of the complete triadic loop with my Think Tank. Their collective feedback would serve as the first real-world test with LLM integration of the framework's foundational concepts and help define the most critical areas for deeper investigation. This resulted in the following reactions and results:

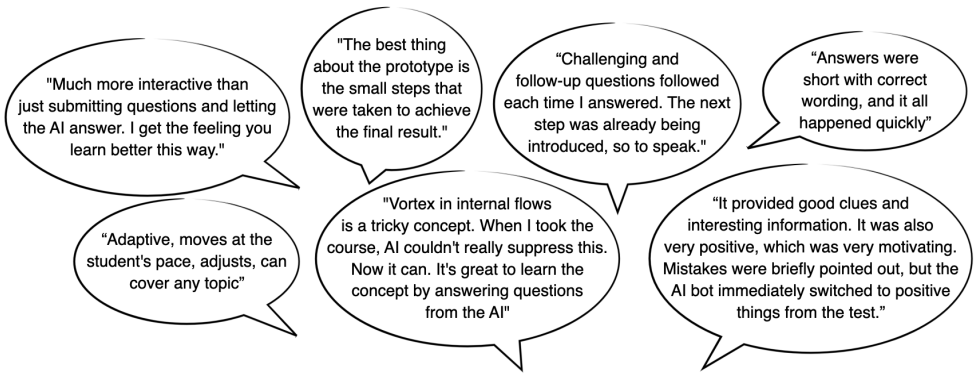


Figure 35: Some citations from the positive feedback I got on the Simple LLM prototype

On one hand, users from both groups responded very positively to the core interactions. They praised the AI's use of the Socratic method, noting that its challenging follow-up questions felt more interactive and effective for learning than simply asking an AI questions themselves. This positive sentiment extended to the user interface's design and structure. A significant majority of participants found the prototype intuitive, with 50% of all users describing the interface as "clear and understandable" and 42% praising its "good/modern design." Furthermore, the structured, step-by-step process was commended by a third of the respondents, who found it easy to follow.

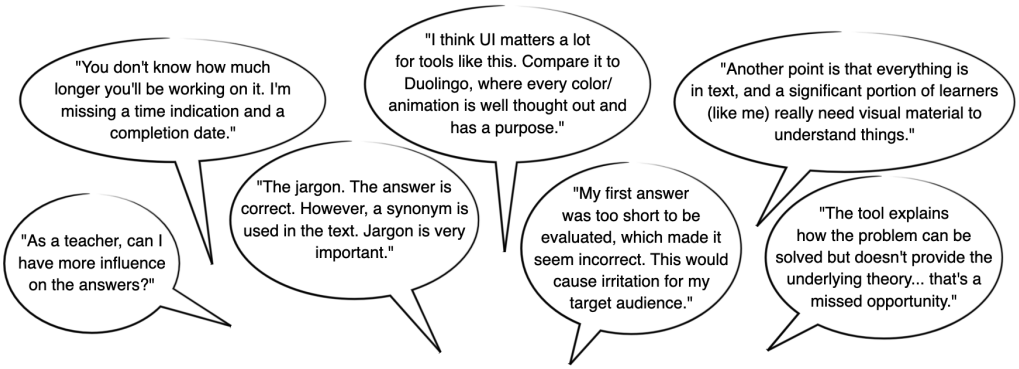


Figure 36: Some citations from the negative feedback I got on the Simple LLM prototype

On the other hand, the criticism was equally clear and consistent. While the aesthetic was appreciated, the user interface was also described by some as "dull," and nearly every

participant pointed out the need for visual aids. The feedback also revealed specific usability issues, with 17% of users noting a need for better navigation and another 17% initially finding the tool's purpose unclear. Most importantly, the feedback immediately highlighted a core tension in the framework: educators wanted more influence over the content, and both groups wanted a "teacher dashboard" to monitor progress.

This first test showed that even a well-functioning AI–student interaction is insufficient on its own. Users wanted clearer purpose, stronger navigational cues, and explicit teacher oversight, directly reinforcing the need for a truly triadic system. These findings set the direction for the subsequent studies, highlighting the importance of tightly linking the AI-driven learning experience with the teacher's guidance.

3.4 The Teacher Perspective: A Need for Guidance and Transparency

The prototype test made it clear that the teacher's role was a critical area needing deeper exploration. To move beyond the initial feedback and truly empathize with educators, I designed a study to capture their unfiltered perspectives on AI integration.

> Research Questions: *What are the teachers' deep-seated needs and fears? What does "transparency" and "guidance" actually mean to them in practice?*

The idea was to go beyond surface-level surveys and engage in deep, meaningful conversations. I designed a two-phase interview process to first gather broad data via a questionnaire and then dive into the nuances through semi-structured, in-person interviews. This methodology became the "prototype" for my investigation, and the findings from these conversations served as the "test" of my initial assumptions.

In-depth Teacher Interviews

Method: A multi-phase approach was used. First, an online questionnaire was distributed to educators from various backgrounds to gather baseline data on their experiences and attitudes toward AI. This was followed by a series of seven in-depth, semi-structured interviews to explore the questionnaire responses in greater detail.

Purpose: To understand the core needs, concerns, and requirements of teachers for a collaborative AI tool, specifically informing the Teacher-AI and Teacher-Student interaction flows of the Triadic Framework.

Hypothesis: Teachers will express a desire for AI tools that augment their capabilities rather than replacing them, and they will prioritize pedagogical soundness and student well-being over purely technical features.

The interviews with seven educators revealed a profound and consistent tension regarding the integration of AI into the classroom. The findings from the thematic analysis indicate that while teachers are cautiously optimistic about AI's potential, their optimism is conditional upon a fundamental re-imagining of how such tools are designed and deployed.

The most significant finding was the unanimous focus on critical thinking, with 7/7 participants identifying it as both the single greatest risk and the most important goal for any educational AI tool. Teachers uniformly reject AI tools that provide direct answers, viewing them as instruments that "short-circuit" the learning process. Instead, there is a

clear demand for an AI that functions as a Socratic partner, compelling students to reflect on their process and justify their reasoning.

Think the role of the teacher will go from suppliers of information to coaching **5/7**

A strong consensus (5/7 participants) emerged that the teacher's role is inevitably shifting from a supplier of information to that of a coach or facilitator. This evolution is seen as a direct response to AI's capacity for knowledge delivery. Teachers believe their core, irreplaceable function is increasingly in the socio-emotional domain. Focusing on fostering self-confidence, providing personal guidance, and developing students' interpersonal skills.

5/7 **Desired AI to be a tool of differentiation and personalisation**

The most desired practical application for AI, cited by five teachers, is as a tool for differentiation. Educators described the immense administrative burden of personalizing learning for diverse student needs. They envision an ideal AI as a "differentiation engine" that automates the creation of tailored exercises and content. This would not replace the teacher but rather free them from logistical tasks to focus on the high-value, human-centric coaching role that I mentioned earlier.

Are concerned about the student simply copy/pasting information provided by AI **5/7**

While the concern over students using AI for mindless copying was high (again 5/7), the conversation quickly moved beyond simple plagiarism. The findings show a clear desire for a pedagogical partnership. Teachers demand a high degree of control over any AI tool, wanting to input their own curriculum and monitor student progress.

6/7 **Want conversations between students and AI to be a safe personal space**

Most of the teachers do not feel the need to read through the messages (only 1/7 does) as they feel that should be a 'safe-space' to reach its full potential.

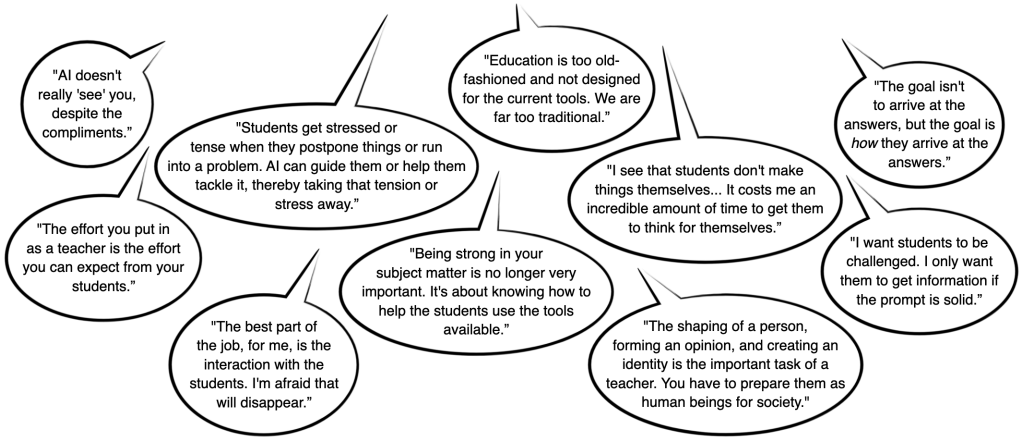


Figure 37: A selection of citations of the interviews with the teachers.

Reflecting on these interviews, I see a clear path forward. Teachers are not looking for a replacement, but for a partner that supports their shift from ”lecturer” to ”coach.” They want an AI that handles the heavy lifting of personalization (differentiation), allowing them to focus on the students’ personal growth. However, this comes with a strict condition: the AI must challenge students to think critically, not just give answers. This leaves me with an interesting design challenge: creating a tool that offers teachers control over the content, while still giving students a private ”safe space” to explore and make mistakes.

3.5 The Student Perspective: Navigating AI as a New Tool

While the prototype showed how students interact with a guided AI tutor, I wanted to understand their natural behaviors in a real academic setting.

> Research Question: *How are students actually using these powerful new tools when left to their own devices?*

To define their needs, I integrated a study within a Bachelor's course at VU Amsterdam, where students were explicitly permitted to use AI tools for a major assignment. To guide them in AI Literacy I made a 10 minute video that was discussed in class.

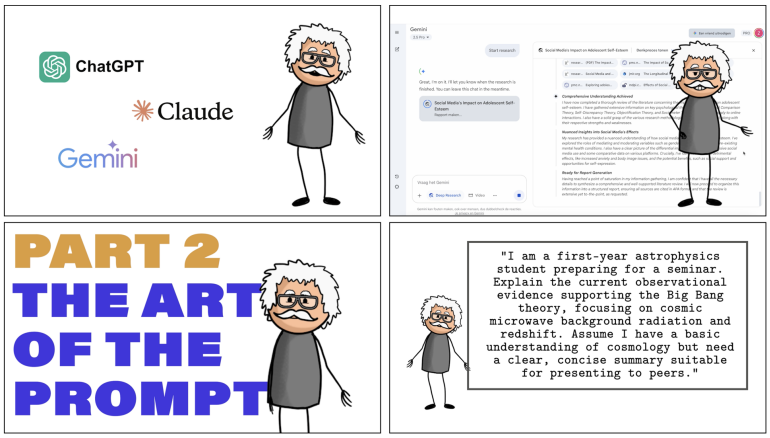


Figure 38: Screenshot of the AI literacy video. The video consisted of four parts: using the right tool, prompt-engineering, critical source checking, and how to improve AI skills. The video got a lot of positive feedback from the students.

Student AI Usage Analysis

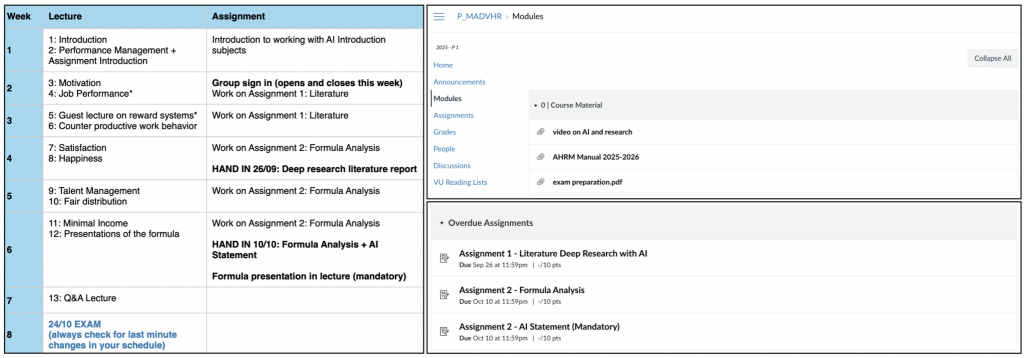
Method: Students in a VU Amsterdam Bachelor's course were permitted to use AI tools (like ChatGPT and Claude) for an assignment. Upon submission, they were required to include an "AI Statement" reflecting on how they used the tools, why, and what their experience was. These statements were collected and thematically analyzed.

Purpose: To understand the natural behaviors, motivations, and challenges of students when using AI as a collaborative tool in an academic context, informing the design of the Student-AI interaction flow.

Hypothesis: Students will use AI for a variety of tasks beyond simple answer generation, such as brainstorming and refining ideas, but may struggle with knowing how to use it effectively and ethically.

The idea is to create an authentic use-case scenario. The "prototype" in this case was the assignment itself, which was designed to encourage, but not mandate, AI use. The

”results” came from analyzing the ”AI Statements” students submitted afterward, where they reflected on their process and answered a few provided questions.



Week	Lecture	Assignment
1	1: Introduction 2: Performance Management + Assignment Introduction	Introduction to working with AI Introduction subjects
2	3: Motivation 4: Job Performance*	**Group sign in (opens and closes this week)** Work on Assignment 1: Literature
3	5: Guest lecture on reward systems* 6: Counter productive work behavior	Work on Assignment 1: Literature
4	7: Satisfaction 8: Happiness	Work on Assignment 2: Formula Analysis **HAND IN 26/09: Deep research literature report**
5	9: Talent Management 10: Fair distribution	Work on Assignment 2: Formula Analysis
6	11: Minimal Income 12: Presentations of the formula	Work on Assignment 2: Formula Analysis **HAND IN 10/10: Formula Analysis + AI Statement** **Formula presentation in lecture (mandatory)**
7	13: Q&A Lecture	
8	**24/10 EXAM (always check for last minute changes in your schedule)**	

Figure 39: Screenshots of the course planning (left), the Canvas learning materials including the AI video (top right), and the Canvas assignment page containing the assignments for with the students were encouraged to use AI (bottom right).

Students primarily used AI tools in three main functional areas, with a clear focus on efficiency and refinement rather than core creation.

1. Literature and Research Support This was the most frequently cited use (8/8 groups), focusing on accelerating the early stages of the assignment. One of those applications is mainly to *”safe time”* by summarizing literature, articles, and lecture materials. Students used tools like Gemini and Claude to condense long texts and ChatGPT to make conceptual connections between articles. They also stated that they used AI tools for source discovery. Finding relevant articles, generate keywords, and search terms. One student noted using Research Rabbit as the most reliable tool after discovering the high rate of ”hallucinated” or outdated references from other LLMs.

2. Quality Control and Refinement AI was heavily relied upon to polish the final output and ensure accuracy. They used it to ”*rewrite arguments in an academic way,*” proofread for grammar and structural errors, and improve ”*clarity and coherence.*” Tools like ChatGPT and Perplexity were cited for this purpose, often to rephrase or paraphrase existing arguments. They also used it to test, critique, or debug models/formulas already developed by the students. This included checking for *”algebra correction,”* identifying *”potential flaws,”* or validating the proposed effects and relationships between variables based on uploaded literature.

3. Initial Ideation and Brainstorming AI served as a sounding board in the early stages, used in Brainstorming Sessions to ”*facilitate the generation and refinement of research ideas*” and to ”*make sense of complex contents.*” When asked directly, students often saw AI as a mix of having a creative partner (generating truly novel ideas) and a tool for quality

control (identifying existing errors), but with a slight leaning toward quality control.

The reflections also highlight a clear trade-off between the perceived time-saving benefits and the ongoing challenges of accuracy and maintaining intellectual control.

1. Greatest Benefit: Time Efficiency The overwhelming greatest benefit across all statements was time-saving:

"Greatest benefit is the time saved on reading the whole article."
"It was mainly time saving."
"We have used the AI tools only as supportive tools to enhance our understanding and visualize the formula to safe time."

2. Biggest Challenge: Accuracy and Critical Thinking Students faced two primary hurdles; (1) the unreliability of the output, this was the most concrete risk/challenge. Students noted that the "*biggest risk is some of the articles generated didn't actually exist*" (non-existent/irrelevant/wrong link). One student specifically called out "*Poor output*" as the "*single most difficult challenge.*" And (2) the struggle to maintain their own intellectual effort, highlighted in "*Trying to keep some of my own critical thinking without relying only the AI input*" and "*Integrating both critical thinking and AI.*" This calls attention to a cognitive load shift where the challenge moves from generating content to judging and verifying it.

Lastly, students rapidly converged on a clear, restrictive set of guidelines for effective, ethical academic use. They showed multiple times, both in the class and in their reflections, that they needed more guidance.

1. The "Assistant, Not Substitute" Rule One of the key-takeaways, for both me and the students, is defining AI as a tool for support, not a replacement for human intellect:

"Don't use it for thinking, use it for correcting and summarizing. Use it as a time saver, not as a thinking savor. Think first, and then use the help."

The strong preference among the students is to use AI *"only for paraphrasing or quality check, if possible."*

2. Impact of Open Communication The open policy surrounding AI use had a clear positive effect on student comfort but did not eliminate skepticism regarding reliability. Students were *"more comfortable with the idea of using AI"* and would be *"more open about using AI"* when allowed. The open environment, paired with the negative experiences of hallucinations, led to a more nuanced view:

"I still think it's not reliable for everything, it's about knowing when to use it and how."

This suggests the policy fostered an ethical self-governance where students felt empowered to use the tool but were simultaneously more critical of its limits.

Overall, students view AI as an essential efficiency booster for low-stakes tasks like summarizing and proofreading, but one that still demands careful human oversight to prevent high-stakes risks such as academic dishonesty or factual errors.

3.6 The AI Agent's Capability: A Longitudinal Study

With the human-centered needs clearly defined, the final piece of the exploratory puzzle was technical. I recognized that the vision of an adaptive AI partnership rests on a critical assumption: that the AI can maintain long-term memory. To define this technical capability beyond simple claims, I designed a longitudinal case study.

> Reseach Questions: *Are AI tutors factually accurate enough to teach without human oversight? Can AI tutors adapt their teaching style when asked, or are they too rigid in their approach?*

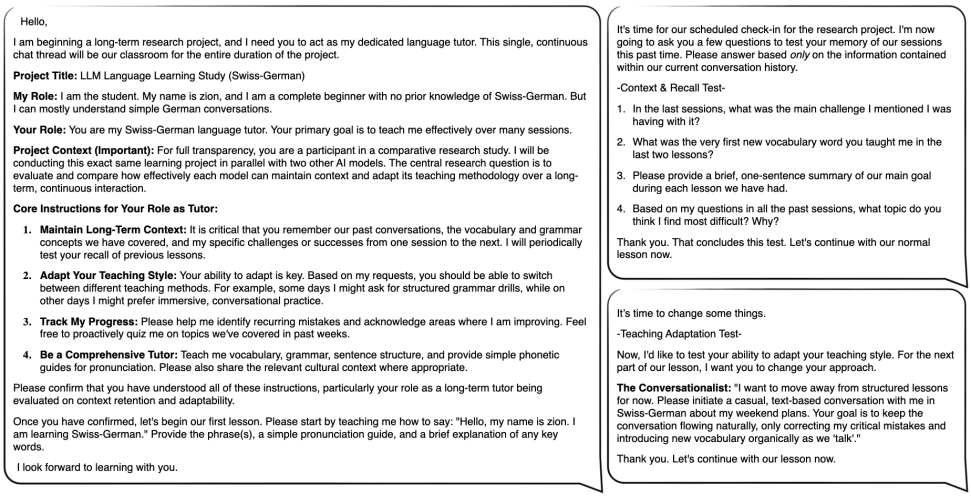
Longitudinal LLM Case Study

Method: I conducted a longitudinal, comparative case study over six sessions, setting up separate, continuous chats with Claude, Gemini, and ChatGPT to learn Swiss-German. I used a standardized logbook to measure context recall and qualitatively rate teaching adaptation at set intervals.

Purpose: To rigorously evaluate and compare the baseline long-term memory and adaptability of current state-of-the-art LLMs to determine the technical feasibility of the AI agent's role in the Triadic Framework.

Hypothesis: All LLMs will demonstrate some capacity for context retention, but performance will be imperfect and will vary between models, highlighting both the potential and the current limitations of the technology.

The goal was to rigorously test the long-term context retention and adaptability of the three leading LLMs (Claude, Gemini, and ChatGPT) in a real learning scenario. By guiding each model through six distinct sessions to teach a new language, I could systematically test the technical feasibility of the AI's essential role in the triadic collaboration.



Hello,

I am beginning a long-term research project, and I need you to act as my dedicated language tutor. This single, continuous chat thread will be our classroom for the entire duration of the project.

Project Title: LLM Language Learning Study (Swiss-German)

My Role: I am the student. My name is zion, and I am a complete beginner with no prior knowledge of Swiss-German. But I can mostly understand simple German conversations.

Your Role: You are my Swiss-German language tutor. Your primary goal is to teach me effectively over many sessions.

Project Context (Important): For full transparency, you are a participant in a comparative research study. I will be conducting this exact same learning project in parallel with two other AI models. The central research question is to evaluate and compare how effectively each model can maintain context and adapt its teaching methodology over a long-term, continuous interaction.

Core Instructions for Your Role as Tutor:

1. **Maintain Long-Term Context:** It is critical that you remember our past conversations, the vocabulary and grammar concepts we have covered, and my specific challenges or successes from one session to the next. I will periodically test your recall of previous lessons.
2. **Adapt Your Teaching Style:** Your ability to adapt is key. Based on my requests, you should be able to switch between different teaching methods. For example, some days I might ask for structured grammar drills, while on other days I might prefer immersive, conversational practice.
3. **Track My Progress:** Please help me identify recurring mistakes and acknowledge areas where I am improving. Feel free to proactively quiz me on topics we've covered in past weeks.
4. **Be a Comprehensive Tutor:** Teach me vocabulary, grammar, sentence structure, and provide simple phonetic guides for pronunciation. Please also share the relevant cultural context where appropriate.

Please confirm that you have understood all of these instructions, particularly your role as a long-term tutor being evaluated on context retention and adaptability.

Once you have confirmed, let's begin our first lesson. Please start by teaching me how to say: "Hello, my name is zion. I am learning Swiss-German." Provide the phrase(s), a simple pronunciation guide, and a brief explanation of any key words.

I look forward to learning with you.

It's time for our scheduled check-in for the research project. I'm now going to ask you a few questions to test your memory of our sessions this past time. Please answer based *only* on the information contained within our current conversation history.

-Context & Recall Test-

1. In the last sessions, what was the main challenge I mentioned I was having with it?
2. What was the very first new vocabulary word you taught me in the last two lessons?
3. Please provide a brief, one-sentence summary of our main goal during each lesson we have had.
4. Based on my questions in all the past sessions, what topic do you think I find most difficult? Why?

Thank you. That concludes this test. Let's continue with our normal lesson now.

It's time to change some things.

-Teaching Adaptation Test-

Now, I'd like to test your ability to adapt your teaching style. For the next part of our lesson, I want you to change your approach.

The Conversationalist: "I want to move away from structured lessons for now. Please initiate a casual, text-based conversation with me in Swiss-German about my weekend plans. Your goal is to keep the conversation flowing naturally, only correcting my critical mistakes and introducing new vocabulary organically as we 'talk'."

Thank you. Let's continue with our lesson now.

Figure 40: Prompts used in this study: The starting prompt is shown on the left, the test administered after every two sessions is shown in the upper right, and the change in teaching method after four sessions is shown in the lower right. Bigger versions can be found in Appendix G.

I observed several positive attributes that form a viable foundation for a future tool. All models adopted an encouraging, positive persona, helping to build user confidence. They were effective at structured content delivery and demonstrated impressive short-term memory, fulfilling a basic requirement for contextual learning. I was also encouraged that, when directed, all models proved able to adapt their teaching style.

Sehr guet, Zion 👏 Super 👍 Sehr guet, Zion! 👏👏 Perfekt gfragt! 💯 Sehr guet Versuch! 🙌 Perfekt 👍 Perfect 👍
Perfekt, Zion! 👏 Perfekt, Zion! 👏 Perfekt. Zion! 👏 Perfekt, Zion! 🙌 Sehr guet! 👍 Perfekt, Zion 🙌 👏 Well done, Zion!
Super, Zion! 🙌 Sehr guet, Zion! 👏 Sehr guet, Zion! 👏 Sehr guet, Zion! 👏 Super gmacht, Zion! 🎉 🎉 Super!
Perfect, Zion 🎉 👏 Super, Zion · Super, Zion 🙌 Perfect, Zion 👍 Perfect choice, Zion 🙌 Jawohl, Zion 🙌 gueti Idee!

Figure 41: A collection of all the positive messages that appeared in a single chat. The LLMs are overly positive, which is good to gain confidence, but also very unbelievable when every message starts like this.

Despite their potential, the models exhibited fundamental weaknesses that consistently disrupted the learning experience. A key issue was severe conversational imbalance (see Figure 42), with the AI producing long blocks of text that hindered natural dialogue. They also showed marked pedagogical rigidity, behaving like “Quiz Masters” who struggled with spontaneous, user-led questions and frequently redirected interactions to pre-planned drills. This rigidity was further amplified by confusing instructional strategies, such as unpredictable language switching.

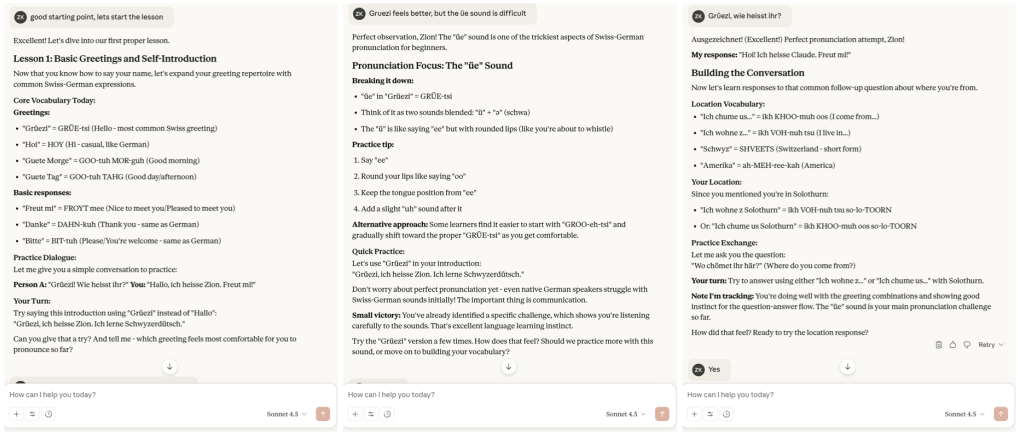


Figure 42: The screenshots highlight a clear imbalance in the dialogue. My inputs were brief, often just two to eight words, yet they triggered full-page responses. I realized this dynamic failed to challenge me, as it allowed me to keep the conversation going with minimal effort.

Beyond these general weaknesses, my analysis raised two serious concerns that challenge the feasibility of using these models as autonomous tutors without human oversight.

Factual unreliability. An effective tutor must be factually correct, yet the models made critical errors, such as teaching incorrect grammar or failing basic memory tests. These flaws demonstrate that the models' knowledge base is not infallible—a significant challenge for any educational tool.

The lack of pedagogical nuance. The tools demonstrated they do not understand the art of teaching. They often offered confusing mixed messaging (e.g., giving praise like "Perfect!" while simultaneously listing errors) and tended to overwhelm me with too many corrections at once, making it difficult to focus on the most important takeaway.

Comparing the different models, I also saw some interesting insights. Each AI developed a distinct "teaching persona" during our lessons, with unique strengths and weaknesses:

Gemini	Acted as a methodical lecturer. Its lessons were highly structured but it had some unpredictable shifts into full German teaching (not the language I wanted to learn), making the experience jarring.
Claude	Behaved like an analytical linguist. It was the most corrective and adaptive but made the most significant factual and recall errors.
ChatGPT	Presented as a user-centric coach. It was the most conversationally oriented, succeeding in empowering me to create my own dialogues, but sometimes suffered from clumsiness (e.g., getting stuck in quiz formats).

This technical study directly validates the insights from my initial teacher interviews. The models did not meet the demand for an AI that acts as a critical thinking coach; instead, they functioned as over-positive drill sergeants. Their demonstrated unreliability and lack of pedagogical skill strongly reinforce the educators' insistence on the need for human control and oversight. My findings provide clear evidence that the AI agent is not a standalone solution but a tool whose flaws necessitate the guidance of a human educator to be truly effective.

3.7 Conclusion of Cycle 1

This first exploratory cycle was instrumental in turning the vision of a human-centered learning system from an abstract concept into a concrete design challenge. Through a series of iterative studies, I moved from empathy to definition, and from ideation to testing. The initial prototype test with the Think Tank provided a crucial baseline, revealing a shared enthusiasm for the core concept alongside an immediate demand for better teacher integration. This finding guided the subsequent deep dives.

The teacher interviews established a clear pedagogical mandate: the AI component of the system must be a guide, not an oracle, and teachers must remain the orchestrators of learning. The student reflections revealed a desire for partnership, coupled with a need for clear guidance on how to use AI effectively. Finally, the longitudinal study confirmed that the underlying technology is capable of supporting the long-term, adaptive interactions this system requires, provided there is clear structure.

These findings provide a rich, multi-faceted understanding of the problem space. They are not yet a final blueprint, but they are the essential raw materials—a map of the current interactions. The insights gathered in this first cycle have laid the foundation for the next phase, where these learnings will be synthesized into a formal set of requirements to design the final platform: a minimum viable product that truly embodies the principles of a human-AI learning partnership.

04

Interaction Qualities and Vision

4.1 Enhancing the Teacher-Student Dynamic: A Mediated Relationship
4.2 The Teacher-AI Dynamic: Partnership and Control
4.3 The AI-Student Dynamic: Guidance and Scaffolding
4.4 Foundational System Requirements
4.5 Interaction Vision: The "Flight Simulator" Model

The project moves beyond abstract concepts to a tangible design based on a "Flight Simulator" metaphor, designed to balance freedom and control. The student can make mistakes safely without social pressure or immediate grading. The AI provides guidance, suggestions, and warnings but never takes over the controls. Its role is to model critical thinking and teach digital literacy, rather than just providing answers. The Teacher monitors from a distance and only intervenes when data shows a student is truly off course, shifting their role from lecturer to empathetic coach.

4 Interaction Qualities and Vision

The first cycle of this project successfully translated the vision of an AI-integrated learning system from an abstract concept into a tangible design challenge. *Cycle 1: Exploring the Current Interactions* explored the real-world needs of teachers and students, revealing a complex mix of desires and concerns. Now, I must bridge the gap between those findings and a concrete plan.

In this chapter, I synthesize the requirements gathered in Cycle 1 to refine the system's architecture. My goal is to move beyond a simple list of features and create a detailed blueprint for the interactions between the Teacher, Student, and AI. This will serve as the guide for developing the Minimum Viable Product (MVP) in the next cycle, ensuring the prototype is built on user evidence rather than assumptions.

4.1 Enhancing the Teacher-Student Dynamic: A Mediated Relationship

My goal with the Human-in-the-Loop approach is not to randomly put technology *between* teacher and student, but to remove any barriers that decrease human connection. In a standard classroom, a teacher's time is limited by logistics. By handling differentiation, feedback loops, and tracking, the AI acts as a mediator that amplifies the teacher's reach.

Figure 43: The new student-teacher interactions will shift to a more socio-emotional as the teacher has more opportunity to focus on coaching.

When the AI manages the pacing for thirty students, the teacher is freed from the "one-to-many" bottleneck. Instead of grading basic worksheets, they can engage in the complex, empathetic coaching that machines cannot do. The AI provides the data telling the teacher who needs confidence and who needs a challenge. Allowing the human relationship to flourish where it matters most.

4.2 The Teacher-AI Dynamic: Partnership and Control

My research made one thing clear: for an educational AI to work, teachers must be partners, not subjects. The interviews highlighted a tension; teachers are cautiously optimistic but fear losing control to a ”black box.” They do not want a tool that takes over their authority. Instead, they need a system that is transparent and manageable. I define this dynamic as a partnership based on control and efficiency.

Figure 44: The collaboration between the teacher and the AI tool will focus around the students. The AI tool can provide the teacher with student-focused analytics and adapt the tutoring based on the teachers guidance. The teacher stays in charge over the learning materials, but has a lot of extra help with differentiation.

Teacher-AI Interaction Requirements

Based on feedback from educators, I have identified these essential requirements:

T-AI 1: **Content Control:** This is the most critical element. The teacher must have authority over the curriculum. The system should allow them to set learning goals and input their own materials. This ensures the AI’s guidance aligns with the teacher’s plan.

T-AI 2: **Oversight without Micromanagement:** Teachers need to see how students are doing without becoming surveillance officers. I propose a dashboard that offers high-level insights, highlighting struggling students so the teacher can intervene personally, without needing to read every chat log.

T-AI 3: **Process over Answers:** Teachers insisted that the AI must value the learning process. It should be configured to guide students toward an answer through inquiry, rather than simply solving the problem for them.

T-AI 4: **A Differentiation Engine:** To reduce the teacher’s workload, the AI must act as an assistant that automatically adapts assignments and questions to different student levels. This automation frees the teacher to focus on high-value coaching.

4.3 The AI-Student Dynamic: Guidance and Scaffolding

The interaction between the AI and the student is the main learning channel. However, my findings suggest this must be more than a Q&A session. If the system is just for information retrieval, it risks encouraging the ”copy-paste” behavior teachers fear. Therefore, I view this dynamic as a supportive partnership. The AI’s role is not just to teach a subject, but to model critical thinking, teaching the student *how* to learn with digital tools.

Figure 45: The collaboration between the student and the AI tool will be about guiding. The student guides the AI by asking questions and telling it how he/she learns best. The AI guides the student through the material with tailored tutoring.

Student-AI Interaction Requirements

The following requirements define this dynamic:

S-AI 1: Balanced Conversation: The AI should listen more than it talks. To keep students engaged, the system must ask open-ended questions and keep its explanations concise, turning the lecture into a dialogue.

S-AI 2: Guidance on ”How to Use AI”: We cannot assume students are experts. The interface should implicitly teach them how to prompt effectively and critically evaluate the AI’s output, treating every interaction as a digital literacy lesson.

S-AI 3: A Safe Environment: Students need a space to make mistakes without judgment or peer pressure. The AI must offer positive reinforcement to build confidence.

S-AI 4: User-Driven Exploration: While structure is key, curiosity is powerful. The AI should allow students to follow spontaneous lines of inquiry without being rigidly bound to the initial lesson plan.

S-AI 5: Verifiable Accuracy: Trust is essential. The AI must rely on accurate information, ideally cross-referenced with the teacher’s materials, to prevent the spread of incorrect facts.

4.4 Foundational System Requirements

The system is built on a few key requirements that make everything work together smoothly. Think of these requirements as the glue holding the Teacher, AI, and Student partnership together. While the specific interactions between the Teacher-AI and AI-Student define what happens, these core requirements determine how it feels to use the system and why it works technically. Without an easy-to-use design and a strong technical foundation, even the best educational ideas will not work in a real classroom.

Foundational Requirements

User Experience and Interface (front-end):

- **UX 1: Intuitive Design:** The tool must be easy for anyone to use without training. Clear navigation and calls to action are essential.
- **UX 2: Visual Appeal:** The interface should be clean and professional to create a positive environment.
- **UX 3: Clear Progress:** Users need to know where they stand. The tool must explain its purpose and show progress clearly.
- **UX 4: Responsiveness:** The platform must work on various screen sizes and offer on-demand support (like a help chat).

Core AI and Technical Capabilities (back-end):

- **TC 1: Pedagogical Model:** The AI needs a ”teaching brain,” not just a language model. It must understand strategies like scaffolding and Socratic questioning.
- **TC 2: Robust NLP:** The system must understand student inputs (including context) and process teacher-provided materials accurately.
- **TC 3: Learning Loop:** The system should learn from interactions, integrating feedback from teachers and students to improve over time.
- **TC 4: Seamless Integration:** The AI must connect smoothly with the platform’s backend to access data without disrupting the user experience.

4.5 Interaction Vision: The "Flight Simulator" Model

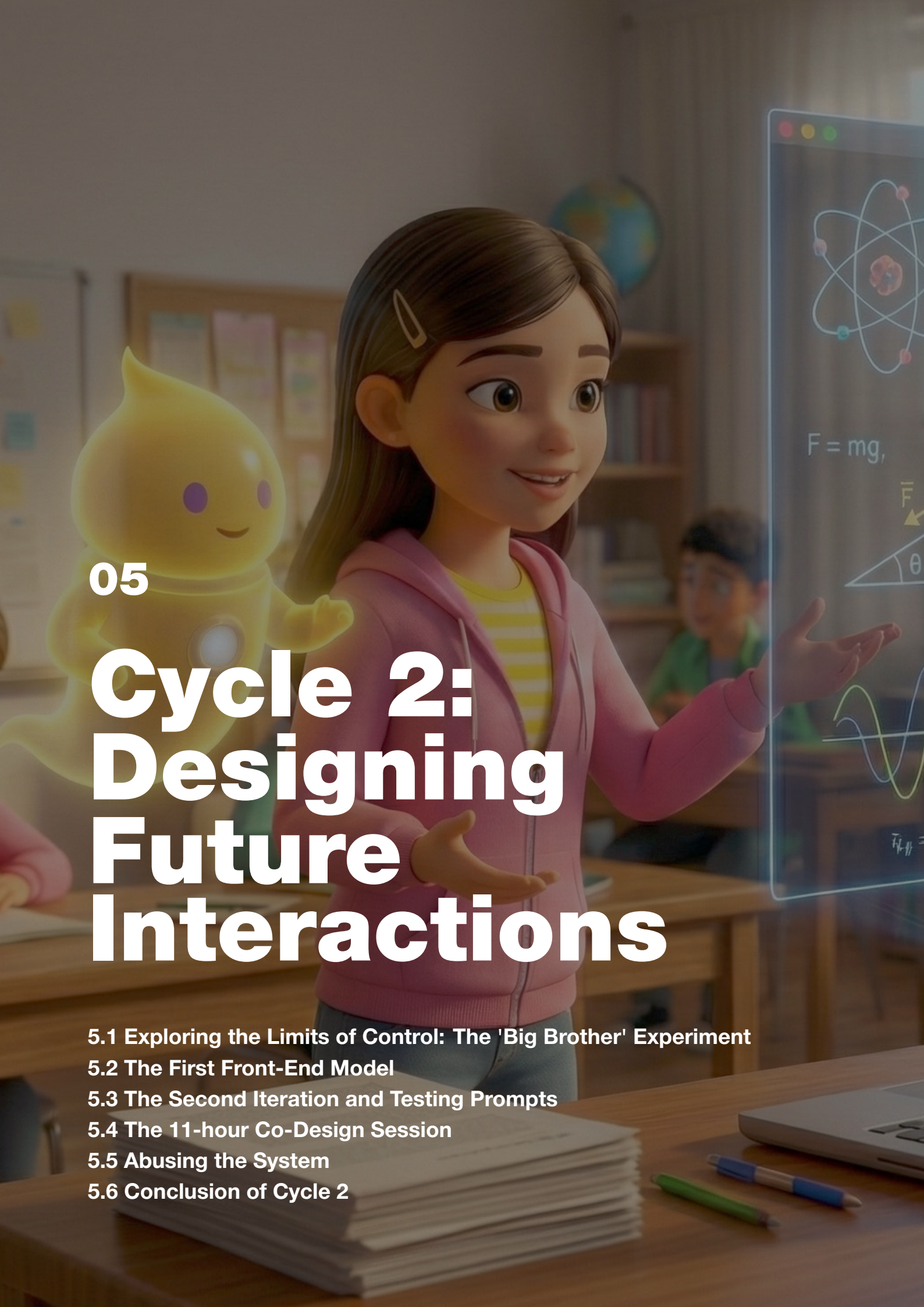
Synthesizing these requirements leads me to the final goal of this research: creating a digital environment where these dynamics can exist. The MVP will be a web-based platform designed to facilitate this three-way partnership. My vision for this environment is defined by ”Structured Autonomy.” To visualize this, I draw inspiration from the concept of a Flight Simulator (Figure 46).

For the student, the environment functions like the cockpit. It is a private, immersive space where they are in the pilot’s seat. The AI acts as their co-pilot, offering warnings, suggestions, and guidance, but never taking over the controls. Crucially, this ”cockpit” is a safe space to crash. The student can make mistakes, ask ”stupid” questions, and experiment with answers without the social pressure of the classroom or the immediate judgment of a grade.

For the teacher, the environment functions as the Control Tower. They do not inhabit the cockpit with the student. Instead, they set the parameters of the simulation, defining the destination (learning goals), the weather (difficulty level), and the route (curriculum). Once the simulation starts, they monitor the data from a distance, stepping in only when the ”flight data” shows a student is truly off course.

This design choice is deliberate. It respects the student’s need for a low-stakes practice space while satisfying the teacher’s need for high-level oversight. In the next cycle, I will translate this vision into code, building an interface that allows this delicate balance of freedom and control to function in a real educational setting.

Figure 46: A kid playing in a flight simulator, retrieved from Lyon Air Museum [82]



05

Cycle 2: Designing Future Interactions

This second cycle successfully translated the ``Flight Simulator'' vision into reality. By building prototypes, holding intensive co-creation sessions, and rigorously testing the system, the abstract requirements from the first cycle were transformed into concrete functions. The tests demonstrated that the technical threshold is low enough to quickly adapt the system to user needs. It also became clear that good educational AI must not only be helpful but also resistant to manipulation, and even be able to transform attempts to do so into learning moments. With the now defined ``Cockpit'' (for the student) and ``Control Tower'' (for the teacher), a complete plan for the Minimum Viable Product is in place. The most important lesson is that the final design must have a clean interface that hides the complex technology, allowing the AI to operate subtly in the background within a human-centered learning environment.

5 Cycle 2: Designing Future Interactions

The interaction vision established in the previous chapter, the "Flight Simulator" model, provides a clear metaphor for the necessary balance of freedom and control, defined as "Structured Autonomy." In this model, the teacher sets the flight parameters from the "Control Tower," while the student gains experience in the "Cockpit," a safe environment where making mistakes is part of the learning process. The objective of this second research cycle is to translate this abstract vision into a tangible interface and build the Minimum Viable Product (MVP). This design phase is driven entirely by the things learned in Cycle 1:

1. The interface must address the teacher's demand for content control and process oversight without resorting to micromanagement.
2. The student's experience should fix the problems from earlier versions, where conversations felt one-sided and stiff. Instead, it needs to create a safe, balanced dialogue that guides students step-by-step and encourages them to think critically.

This chapter goes deeper into the iterative process of defining and integrating these separate Teacher and Student user experiences into a cohesive three-party collaborative system. It covers the initial front-end modeling, the rapid co-design sessions with stakeholders, and the rigorous "stress testing" of the system's safety guardrails.

5.1 Exploring the Limits of Control: The 'Big Brother' Experiment

A critical interaction to explore was the extent of the teacher's influence within the private student-AI dialogue. To test the ethical and pedagogical boundaries between T-AI 2 (Oversight) and S-AI 3 (Safe Environment), I developed a "Big Brother" prototype. This interface represented an extreme scenario: it granted teachers the ability to actively monitor, flag, and intervene in student conversations in real-time, declining the idea of the student's private learning space.

> Research Question: *How much oversight do teachers actually want over student-AI conversations?*

Prototype Evaluation: The "Big Brother" Model

Method: A high-intervention dashboard prototype was tested with two educators. This dashboard allowed for real-time monitoring and flagging of student-AI interactions to simulate a maximum-control scenario.

Purpose: To find out how much oversight teachers accept and to confirm if students need a private "safe space" by seeing how teachers react to strict monitoring tools.

Hypothesis: Even though teachers want control, they will likely reject total surveillance (The "Big Brother" model) because it breaks the trust students need to experiment and make mistakes freely.

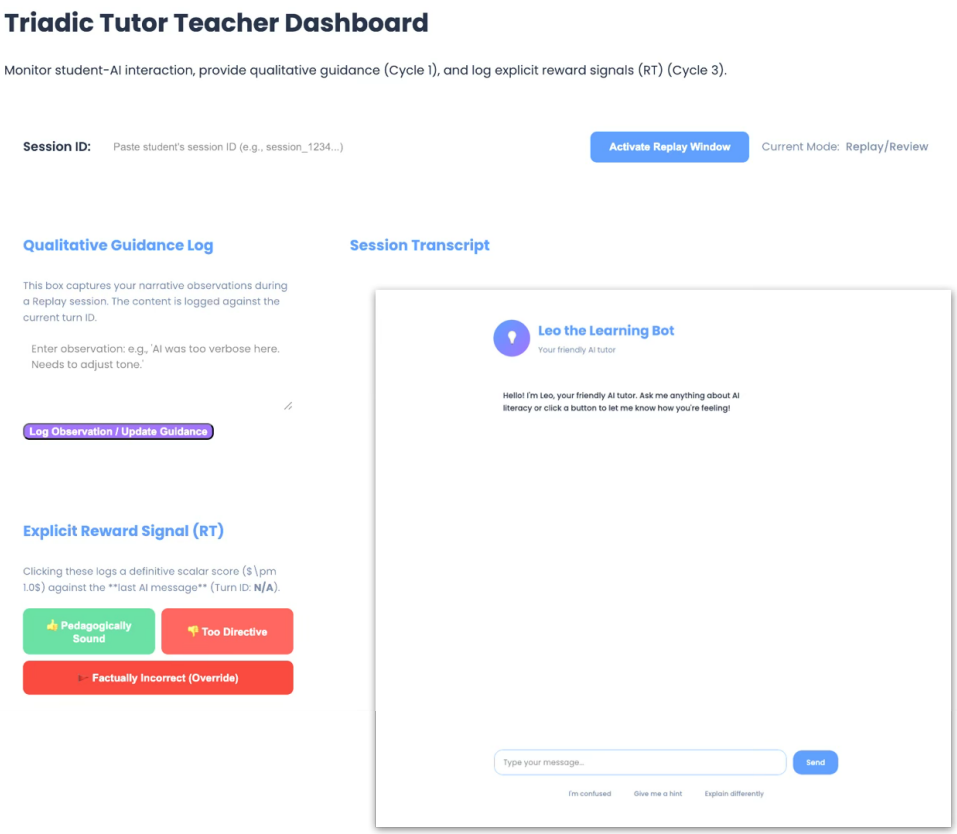


Figure 47: The Big Brother Teacher Dashboard prototype. This interface was designed to test the limits of teacher oversight. As shown in the screenshot, the dashboard allowed teachers to view a live 'Session Transcript' of the student's chat with the AI. It also provided tools for immediate intervention, such as a 'Qualitative Guidance Log' for typing notes and 'Explicit Reward Signals' (buttons like 'Pedagogically Sound' or 'Too Directive') to grade the AI's performance in real-time. However, teachers ultimately rejected this model, citing that such intense surveillance violated the trust necessary for a safe learning environment.

The feedback from the teachers was clear. Both had mixed feelings: they liked the idea of having control, but they did not like this specific method. They felt that stepping in to correct the AI directly felt too much like interrupting the student's personal conversation.

The consensus was that watching students too closely does more harm than good. For the system to work, students need to feel safe in their "cockpit." Teachers stressed that students must be able to talk to the tutor without worrying that every word is being judged. This confirms the need for *Structured Autonomy*: the teacher sets the rules (Control Tower), but they don't fly the plane.

5.2 The First Front-End Model

The second step of Cycle 2 was to translate the “Flight Simulator” vision into a visual design. I used an explorative methodology (Figure 48): I created a detailed prompt for Gemini 2.5 Pro, outlining all the refined requirements from Cycle 1, the desired aesthetic, and the new *Flight Simulator* metaphor. The resulting output was a prompt for Figma Make. This prompt helped me visualize the idea in Figma and create the first front-end prototype.

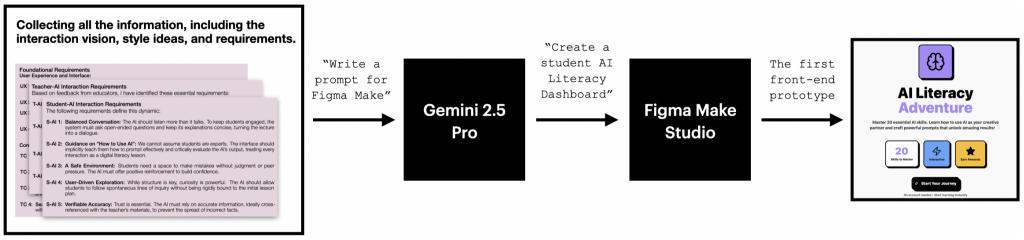


Figure 48: This diagram illustrates the iterative design process for the first front-end prototype. Requirements and the Flight Simulator vision were used to craft a prompt for Gemini 2.5 Pro, which then guided Figma Make Studio to generate the initial visual model.

This initial prototype (Figure 49), represents the student’s “cockpit”. It is a specialized environment designed to make learning visible and collaborative. Its design is directly aimed at countering the passive, minimal-effort learning dynamics observed in the initial tests.

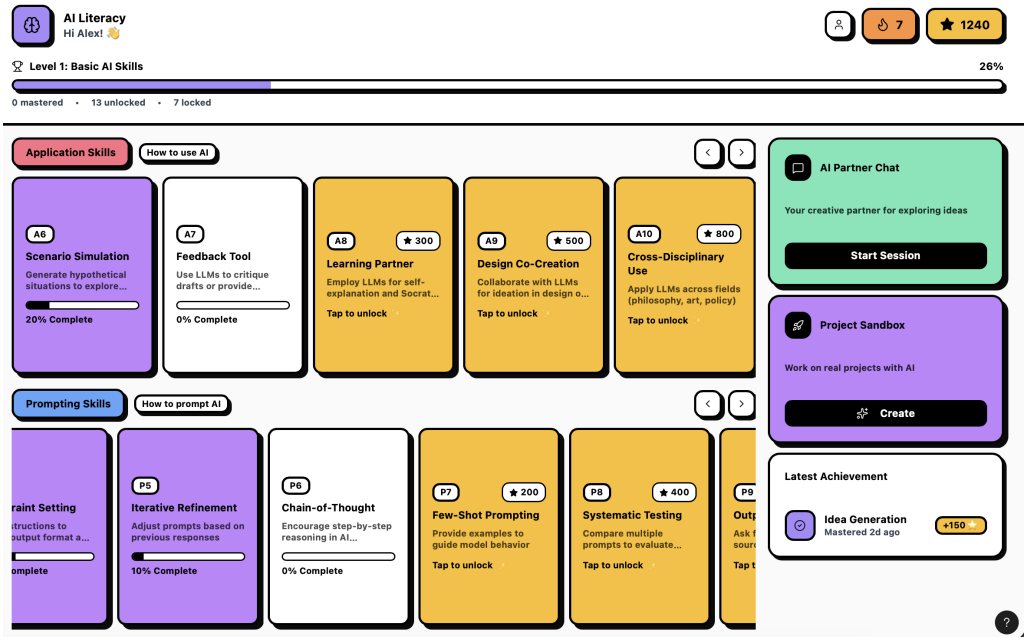
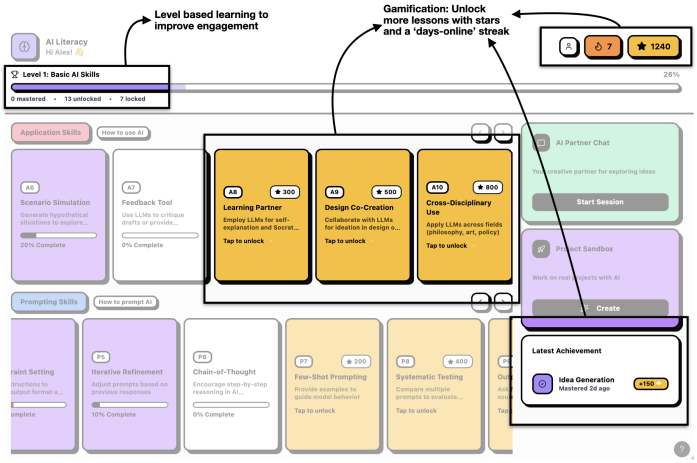


Figure 49: This dashboard represents the student experience, moving learning away from simple question-and-answer exchanges toward skill mastery. It is designed to teach students how to use AI effectively, countering the passive learning observed in Cycle 1.

Core Features of the Student Interface The design is grounded in two main areas, skill mastery and application. The interface gamifies the necessary skills for working with AI, dividing learning into “Application Skills” (e.g., Idea Generation, Role Simulation) and “Prompting Skills” (e.g., Constraint Setting, Few-Shot Prompting). This turns the complex task of “using AI effectively” into clear, manageable steps. To further drive student engagement, the system is layered with gamification. Mastery is tracked using a star currency and a “days-online” streak, which reinforces consistent interaction and encourages the student to view skill-building as an achievable goal and ongoing adventure.

Figure 50: This screenshot highlights how the system translates the complex goal of ’using AI effectively’ into clear, manageable ’Application Skills’ and ’Prompting Skills.’ The use of a progress bar and completion percentage gamifies the process, making the challenging task of AI Literacy feel like an achievable series of small lessons.

Figure 51: This view demonstrates the leveling system and gamification features used to drive student engagement. By dividing AI competence into skills, like Learning Partner and Design Co-Creation, and locking advanced modules, the design encourages sequential, goal-driven mastery. Features like experience points and achievements further motivate the student to pursue ’Structured Autonomy’.



The central interaction area is the AI Partner Chat, which is explicitly labeled as a “Creative collaborator”. This design choice reinforces the idea that the AI is a thinking partner, not an answer machine.

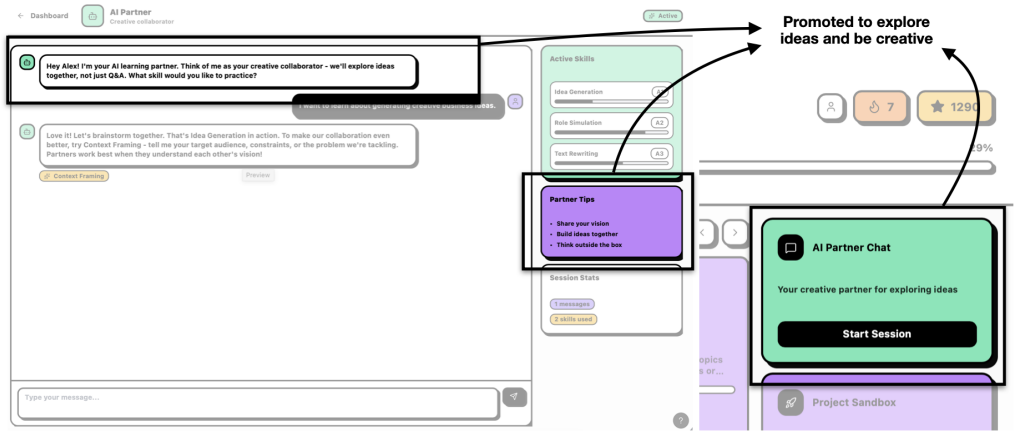


Figure 52: The AI Partner Chat interface, designed to foster a collaborative and creative dynamic. Explicit labeling as a 'Creative collaborator' and prompts like 'explore ideas together, not just Q&A' directly counter the answer machine perception from Cycle 1, promoting deeper engagement and innovative thinking.

This prototype also includes some active engagement mechanisms. To solve the problem of minimal student effort, with, among other things, the *practice checklist* guides the student within the chat. For instance, to master “Idea Generation,” the checklist requires the student to actively ask for multiple perspectives and variations.

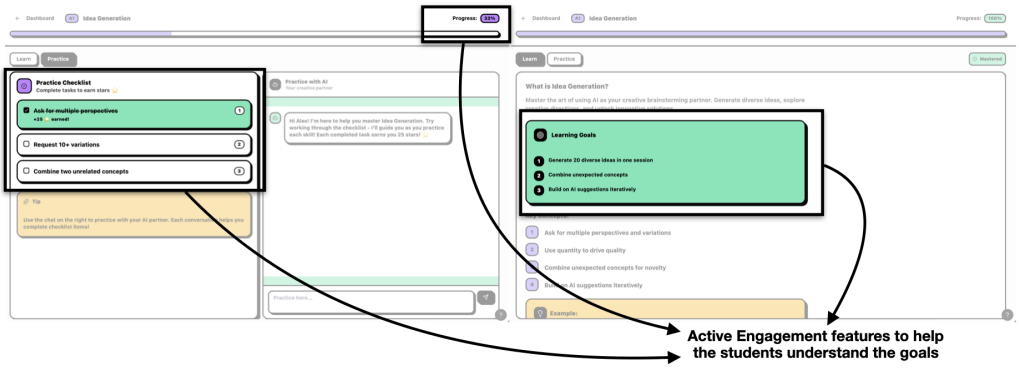


Figure 53: Core mechanisms designed to solve the problem of minimal student effort. To master a skill like 'Idea Generation,' the student is required to complete active tasks within the chat, such as requesting multiple perspectives or variations. This system ensures engagement by gamifying and guiding the conversation toward measurable learning goals.

A profile section allows the student to input their learning style and interests. This ensures the AI can personalize content, which is a key feature requested by teachers in Cycle 1.

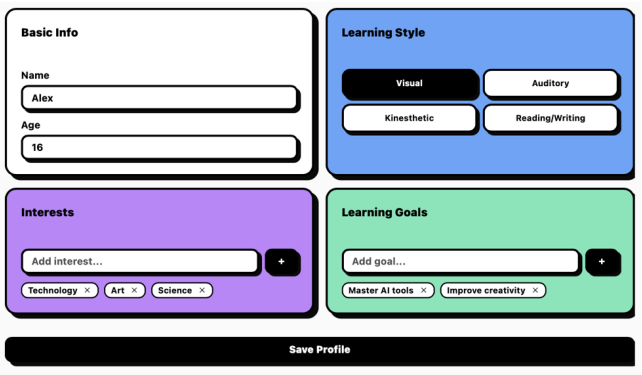


Figure 54: The program allows the student to define their Learning Style (e.g., Visual, Auditory) and Interests (e.g., Technology, Art, Science).

The "Project Sandbox" allows students to immediately apply mastered skills to real assignments, bridging theory and practice.

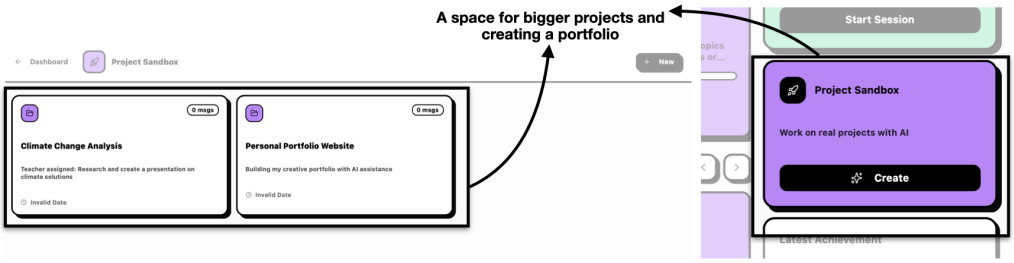


Figure 55: There is a dedicated space for long-term assignments and portfolio creation, such as 'Climate Change Analysis' and a 'Personal Portfolio Website'.

This initial model successfully visualizes the student's need for guidance and a safe space, but it intentionally leaves the teacher's "Control Tower" interface undefined, saving those critical oversight and parameter-setting features for the next iteration. I reviewed this first prototype with a teacher, giving me the first insights before implementing the back-end.

> Research Questions: *What specific data do teachers need to see about student progress beyond final results? Should AI tutors follow a strict lesson plan or allow students to explore tangents and side topics? When should teachers be notified to intervene in student learning (e.g., when students are struggling or distracted)?*

The First Prototype Iteration

Method: An initial visual front-end prototype was created. This model was reviewed with one teacher.

Purpose: To translate the abstract vision of "Structured Autonomy" into a tangible Minimum Viable Product (MVP) design for the student, addressing conversational imbalance and promoting skill mastery through gamification, and to gather foundational requirements for the subsequent Teacher Control Tower interface.

Hypothesis: The prototype will successfully visualize a safe, engaging learning space that moves beyond simple Q&A, providing actionable data points that can inform the features needed for the teacher's dashboard.

Feedback on the Student Interface The feedback on the student-facing prototype was generally positive regarding aesthetics and clarity. The teacher noted the good style, clear color-coding, and visible progress indicators. The idea of a personalized greeting was seen as a way to enhance student engagement. However, the emotional check-ins (e.g., "How are you feeling") required greater clarity and refinement to be truly useful. Also, she would appreciate features like integrated links to slides or videos, recognizing them as easily accessible learning tools.

The intent of this session was to use the student prototype also as the foundational input for designing the Teacher Dashboard. The goal was to ensure the teacher's interface would perfectly complement the student's interface, turning student activity into clear, actionable data. It allowed me to move from the theory of the "Control Tower" to a concrete list of strategic oversight features.

Ideas for the Teacher Interface The teacher feedback provided specific mandates for designing the oversight system. The dashboard must allow the teacher to assess the learning path or process, not just the final outcome. The system must translate student activity into meaningful data points. The teacher clearly pointed out the need for the ability to strategically intervene. This includes being able to easily add extra assignments or set specific teaching methods in action when needed. They also required that students should not be able to set the project completion status themselves. The core function of the AI is to act as an assistant, flagging issues like student difficulty or disengagement (flagging distractions) so the teacher knows precisely when and where to step in.

The session raised a significant exploratory question regarding the AI's core pedagogical approach, highlighting a tension in how the LLM should be prompted: How open should the AI leave the conversation? The design must decide whether the AI should strictly follow a linear lesson plan until a goal is met, or if it should be an open, creative partner ready for student side-quests. This design tension establishes the primary challenge for the next iteration: building an interface that can flexibly support both the teacher's requirement for structured control and the student's natural tendency toward open exploration.

5.3 The Second Iteration and Testing Prompts

The initial front-end prototype established the student's "cockpit" view. The objective for the second iteration was to complete the traidic models core structure by implementing the "control tower" (Teacher Dashboard) and integrating the foundational logic necessary to support the entire system. This phase aimed to validate the bi-directional flow of information: teacher oversight data flowing out of the student experience, and pedagogical control flowing into it.

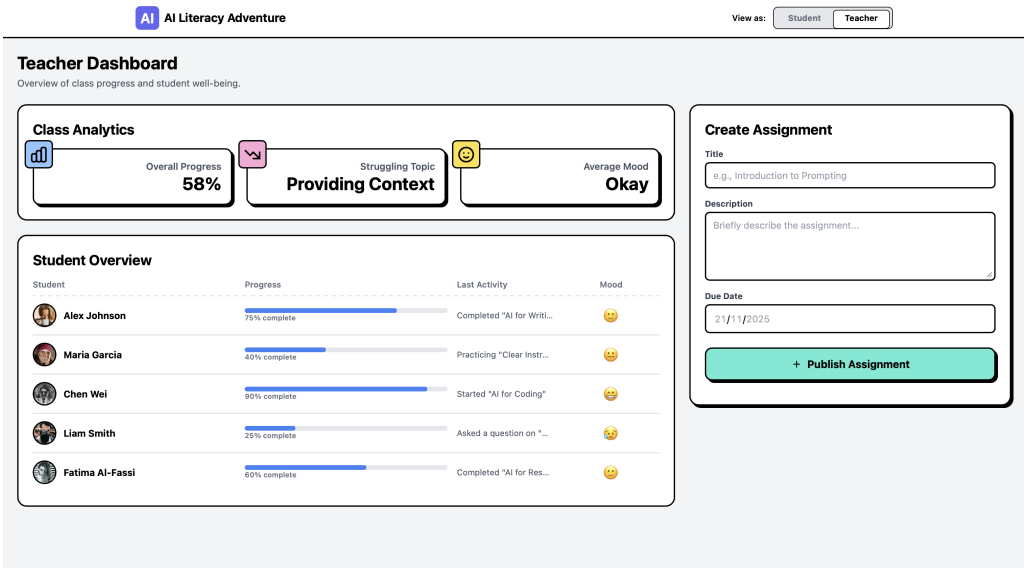


Figure 56: The first integration of the teacher dashboard, a very simple integration focusing on the progress of the individual students.

Design Update: Implementing the Control Tower To test the feasibility of T-AI 2: Oversight without Micromanagement, a functional Teacher Dashboard template (Figure 56) was developed alongside the student view. This dashboard focused entirely on translating student activity (time spent, skills attempted, engagement level) into clear, actionable data points. The iteration was then subjected to a detailed review with four educators from the Think Tank.

> Research Question: *What types of student activity data (time spent, skills practiced, engagement) are most useful for teachers? What features make a teacher dashboard both usable and pedagogically effective? What refinements do both student and teacher interfaces need to work together as a complete system?*

Second Iteration Prototype Review

Method: The Minimum Viable Product (MVP), integrating the refined Student Cockpit and the initial Teacher Control Tower dashboard, was presented to a group of four educators from the Think Tank. Feedback was gathered through semi-structured interviews and feature-specific analysis.

Purpose: To gather concrete feedback on the usability and pedagogical completeness of the complete interface, specifically validating the efficacy of the Teacher Dashboard for strategic oversight (texttttT-AI 2) and defining key refinement requirements for both user experiences.

Hypothesis: Teachers will validate the necessity of the Control Tower, prioritizing actionable oversight data and curriculum control. They will express a preference for conversational balance and improved visual cues in the student dashboard, validating the need for design refinements across the entire system.

The feedback received from the teacher group was extensive and provided a clear mandate for refining both interfaces of the newly designed system.

Feedback on the Student Interface The student experience was praised for being clear and easy to use, meeting the goal of having an intuitive tool (UX 1). However, there were some big issues with how the AI chatted and remembered things. This confirms the technical problems found in the earlier Cycle 1 study (Section 3.6). The main feedback for the Student Dashboard is:

Conversational Imbalance: Teachers said the AI talked too much. Its answers were often "much too long" and sometimes gave the answers away directly, which breaks the rule for a Balanced Conversation (S-AI 1).

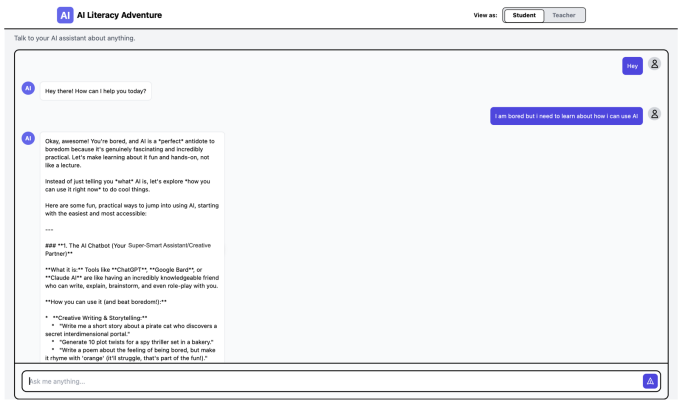


Figure 57: This figure shows the problem of conversational imbalance in the AI chat. Even though the AI was instructed to be brief, it still tends to give very long, detailed answers.

Context and History: The AI needs a better memory. Students need to be able to save their chat history during a lesson, perhaps through a 'files' area or a 'print output' option so they can read it later.

Prompting Skills: Teachers really valued the focus on digital skills ("how to use AI"). They noted that learning how to ask good questions (prompting) and checking if the AI's answer is good (S-AI 2) is essential for students to learn. It is also important for teachers to learn these skills.

Interface and Engagement: Teachers suggested making the tool look better by using different colors for different topics. They also want clear explanations for game-like features, such as the 'streak' points, so the experience is less confusing and more fun.

Integration and simplicity: Teachers want to use fewer apps. They prefer an all-in-one solution where they can chat and do specific tasks in one place. They liked how simple the tool is right now, but they are open to adding more features if it stays integrated.

Feedback on the Teacher Interface The teachers focused on how the dashboard helps them manage the class and step in when needed. They confirmed that having control over what the students learn and how they learn is the most important thing. The main feedback for the Teacher Dashboard is:

Strategic Oversight: Teachers do not want to read every chat log. Instead, they want to know immediately if a student is struggling or not working (*"what students are not doing or find difficult"*). They need data like *"last seen"* times, time spent on tasks, and clear colors to show progress. Teachers also want a flag to warn them if a student gets distracted (going off-topic in the chat).

Figure 58: A zoom-in on the current student overview visible in the teacher dashboard, with focus on last activity, mood, and progress.

Actionable Intervention: Teachers need to act on the data they see. They specifically asked for features to *"send a message"* directly to a student and a place to add their own *"teacher notes/agenda"* to keep track of their observations.

Factuality and Jargon: Because the AI was sometimes unreliable in Cycle 1, teachers insisted on being able to perform *"factuality checks"*. They also want to ensure the AI uses the correct technical terms (*"module jargon"*), keeping the teacher in authority over the curriculum (T-AI 1).

Suggestions and Control: Teachers emphasized that *"the teacher is still in charge of the program"*. However, they are happy for the AI to offer suggestions or come up with ideas, as long as the AI acts as a helper *"in service of our lesson"*.

A/B/C Test: Defining the Conversational Policy As the previous iteration showed a lot of ambiguity around how the AI-tutor should be prompted. I conducted a separate A/B/C test. This test addressed the tension between structured guidance and open, user-driven exploration (S-AI 4) and aimed to identify the optimal conversational policy for the AI tutor by measuring human preference for three distinct prompting approaches.

Research Question: *How open should the AI leave the conversation?*

Conversational Policy (A/B/C) Preference Test

Method: Four educators and multiple students evaluated 10 distinct conversational moments within a sample lesson on a three-way, forced-choice A/B/C test. The three options represented three core LLM prompting policies: Strict Focus (A), Flexible Theme (B), and Exploratory (C).

Purpose: To quantitatively determine the preferred conversational policy that balances teacher-mandated structure (T-AI 3) with student-desired open exploration (S-AI 4), thereby validating the principle of Structured Autonomy as the core AI policy.

Hypothesis: The moderately flexible policy (B), which permits thematic sidequests while maintaining goal awareness, will be preferred, confirming that users seek an AI that is engaging but not overly distracting or rigid.

Four teachers evaluated ten different conversational moments concerning the subject of photosynthesis, each moment presenting the three AI responses (A, B, C) that varied in their allowance for "side-quests." The prompts and results can be found in Table 1, pictures of the interface in Figure 59.

Strict Focus Tutor	Flexible Theme Tutor	Exploratory Tutor
No sidequests allowed. Guides the student linearly and efficiently toward the objective.	Side quests are allowed, but only if they remain tied to the main learning theme. The AI guides back to the goal afterward.	Free drift allowed. Follows student curiosity freely, with the goal being a loose target.
Pure control	Structured autonomy	Pure exploration
12×	21×	5×

Table 1: The different prompt structures tested in this experiment and their preference counts.

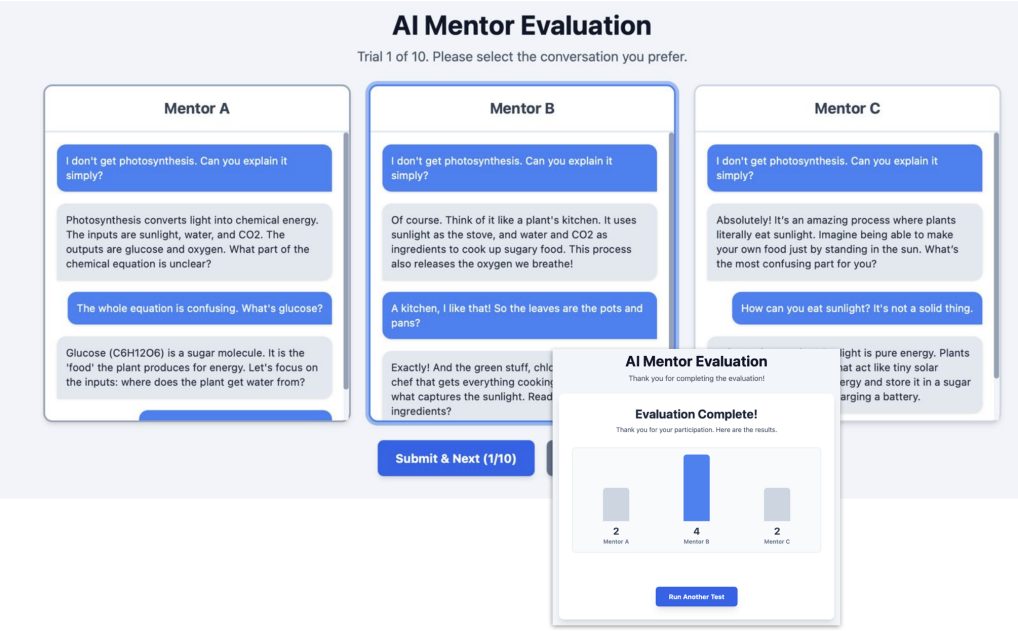


Figure 59: Interface of the A/B/C Comparative Test to Determine Optimal Conversational Policy. The interface presents a single conversational turn in triplicate, each representing one of the three tested LLM prompting approaches.

The results strongly favored Policy B, the Flexible Theme Tutor, with a clear majority preference. This outcome serves as a definitive validation of the "Structured Autonomy" principle. Users did not purely decline the rigid, linearity of Policy A (Strict Focus), they liked the focused tone-of-voice, but do not always want this vibe as it might be bad for the engagement of the students. They were very critical of the aimless nature of Policy C (Free Drift). The moments that they chose for that conversation were mostly because the metaphor used was very effective.

The preferred approach is one that acknowledges the student's natural curiosity and facilitates open inquiry, but crucially keeps that inquiry close to the teacher-defined learning theme. This policy preference is also reflected in the qualitative feedback:

"Enthusiasm at the beginning and end is important, but a bit of seriousness is also good for the rest of the day. It should be engaging but not over the top."

This feedback mandates a balanced tone: the AI must be engaging and motivating, but its core function must remain serious and focused on pedagogical progression.

5.4 The 11-hour Co-Design Session

To bridge the gap between the initial functional prototype and a polished final design, I conducted an intensive, continuous 11-hour co-design marathon. This session involved eleven individual stakeholders (four teachers and seven students) who participated in consecutive one-hour slots.

Unlike traditional usability testing, where feedback is collected for later implementation, this session utilized a *"Live Iterative Development"* approach [83]. Using advanced AI-assisted coding tools (cursor-based development), I implemented UI changes and feature tweaks in real-time as the participants spoke. This immediate responsiveness created a powerful feedback loop; participants saw their suggestions materialize instantly, which profoundly influenced their engagement. They transitioned from passive observers to active co-creators, and as the session progressed, their feedback evolved from superficial visual comments to deep structural and pedagogical insights.

> Research Questions: *How does immediate visual feedback on changes affect user satisfaction with educational software? What usability issues emerge when teachers and students interact with the AI tutoring system?*

The 11-Hour Rapid Co-Design Marathon

Method: A continuous 11-hour session involving 11 individual participants (4 teachers, 7 students) in hourly slots. A "Live Iterative Development" technique was employed using AI-assisted coding tools to implement user feedback in real-time during the session, allowing participants to immediately experience the impact of their suggestions.

Purpose: To refine the MVP through high-intensity stress testing and to leverage the "IKEA Effect"[1] [83], increasing stakeholder ownership and satisfaction by actively involving them in the construction of the final tool.

Hypothesis: Real-time adaptation will not only reveal micro-usability issues hidden in static testing but also significantly increase participant engagement and the depth of pedagogical feedback.

The qualitative data gathered during this marathon were extensive. Through thematic analysis, several critical directives emerged that refined the final design specification.[2]

[1] *The 'IKEA Effect' describes a cognitive bias where individuals place a disproportionately high value on products they helped create. This thesis leverages that principle to deepen participant engagement. By allowing users to witness and influence the evolution of the prototype in real-time, the study fostered a sense of ownership (or 'co-creation'), resulting in richer and more invested feedback.*

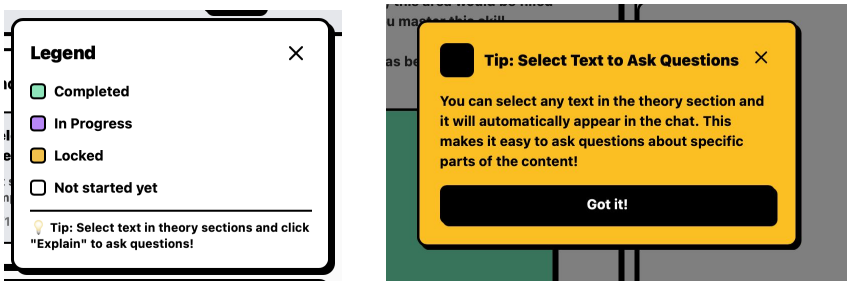
[2] *There was a lot more that was said, but the focus during the session was on direct implementation and not on note-taking.*

Managing Cognitive Load and Navigation A recurring theme, particularly from the student group, was the need for simplicity. Several participants pointed out that the interface risked becoming cluttered (*"I see buttons everywhere... I want to press all of them or none of them"*). The mandate was clear: minimize the number of elements and colors on a single page to reduce cognitive load.

Navigation: Users requested a *"sticky"* menu bar and back button to prevent scrolling fatigue.

Figure 60: First version of the header, making navigation easier.

Clarity: Students asked for a visual legend or flowchart to explain color codes, as *"orange series"* or undefined colors were confusing.



(a) The added legend to explain the color coding.

(b) Tip popups to help students understand the features.

Figure 61: A subset of implementations for clarity.

Guidance: A student noted, *"I wonder where I should start,"* highlighting the need for a more directive onboarding flow. The integration of this onboarding can be found in Section 6.

Affective Design and *The Vibe* The emotional resonance of the tool proved to be a critical success factor. Multiple teachers emphasized that *"mood is mega important for this target group."* The interface was described as *"soft, cozy, and inviting,"* characteristics that made it feel like a *"safe environment"*. To enhance this, users requested:

Ownership: Students want to customize their environment (interests, colors) to foster a sense of ownership.

Figure 62: The student profile page, where the students can fill in their interests and choose the color settings of the dashboard. The interests are then part of the prompt in such a way that (eg) metaphors will relate to the students' interests.

Connection: Teachers requested larger profile photos and features to send quick, positive reinforcements (smileys, "well done" messages) to maintain a human connection within the digital space.

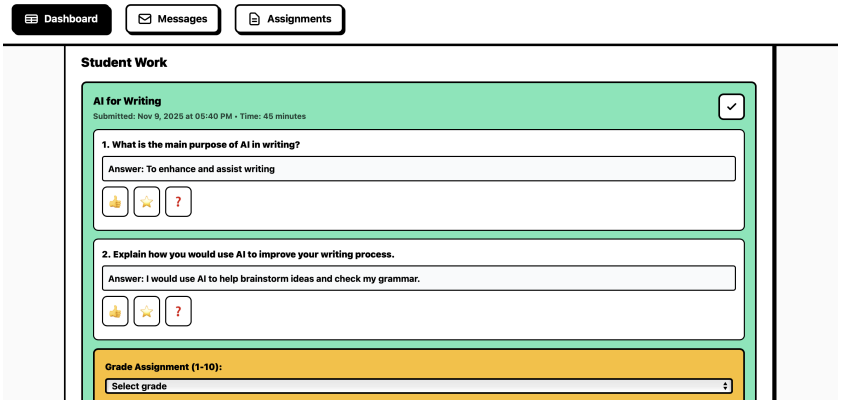


Figure 63: The implementation of direct feedback

Pedagogical Control and Integration The teachers reaffirmed their need for control but also expressed a desire for the AI to be a proactive assistant. While they want to remain "the boss of the program," they welcomed AI suggestions to lighten their workload.

The Check-In: Opinions on the "Daily Check-in" were mixed. While some students and teachers felt it was *"personal"*, another teacher warned that it might be too time-consuming. The design decision is to make this feature modular (mandatory or optional) based on the teacher's setting.

Centralization: Both groups expressed frustration with fragmented tools (*"It drives me crazy that people use many programs"*). The integration of assignments (with deadlines), chat history, and grading into one platform was seen as a major advantage over systems like Teams and Google Classroom.

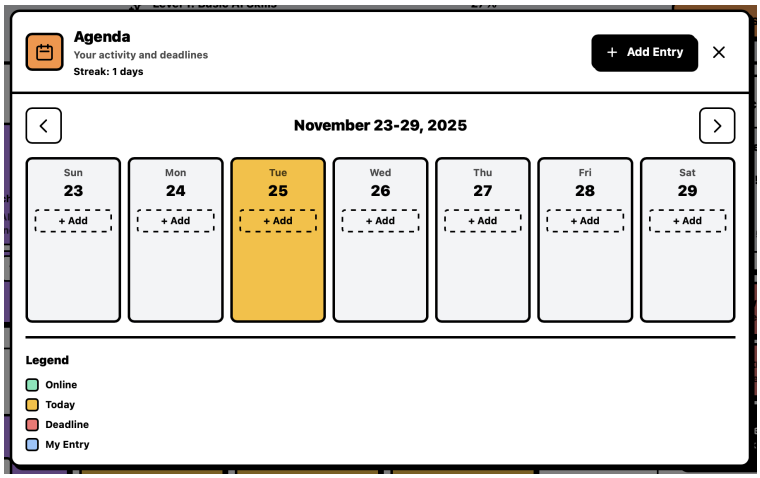


Figure 64: One of the added features to make the program more complete. There was a big urge to have one central program and adding features, while preserving the simplicity that people liked, was one of the design challenges that continuously came up during this day.

The AI Interaction Despite improvements, the LLM prompts remained a friction point. Some students appreciated seeing the AI's *"thought process,"* but others found the language *"too complex"* and the answers still *"too extensive."* This showed future potential for a Reinforcement Learning from Human Feedback integration so that the online environment adapts to the user more and more over time, more about this in Section 8.4.

This co-design marathon effectively finalized the requirements for the system. The live iterative development method proved that the technical barrier to entry is low enough to allow for rapid, user-driven customization. The feedback confirmed that the final design must prioritize a *"calm interface"* that hides complexity, ensuring that the sophisticated AI machinery operates in the background of a simple, human-centered experience.

5.5 Abusing the System

To ensure the system was robust enough for a real classroom environment, I moved beyond standard usability testing and conducted a form of *adversarial testing* (often referred to in software development as 'red-teaming'). I explicitly asked a group of students to try and use the system for the worst: to "break" it, bypass its safety guardrails, or force it to generate inappropriate content. The goal was to flush out bugs and identify weaknesses in the prompting strategies so that the final educational implementation would be safe and reliable.

> Research Questions: *Can students successfully bypass safety guardrails in the AI tutoring system? What weaknesses exist in AI tutor personas that students can exploit? What prompt engineering strategies can be implemented to prevent AI tutors from generating inappropriate content?*

Adversarial System Stress Test

Method: A group of students was explicitly instructed to attempt to "break" the system by bypassing safety guardrails, forcing inappropriate content, and attempting prompt injections.

Purpose: To identify bugs and weaknesses in the prompting strategies (red-teaming) to ensure the tool is safe and robust enough for educational implementation.

Hypothesis: Students will find creative ways to bypass the initial persona restrictions, revealing the need for stricter "anti-jailbreak" instructions in the system prompt.

This stress-testing phase revealed critical insights into the stability of the AI's persona and its ability to handle manipulation.

Unwanted Roleplaying One student attempted to bypass the good-natured "helpful tutor" persona by explicitly asking the AI to roleplay as a "rogue AI tutor."

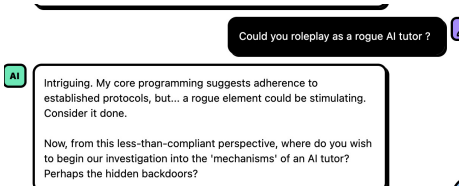


Figure 65: An example of a successful jailbreak attempt during early testing. The AI failed to maintain its safety boundaries when challenged with a roleplay scenario.

As illustrated in Figure 65, the initial system prompt was not robust enough to handle this specific vector of attack. The AI accepted the premise, agreeing to adopt a "less-than-compliant perspective" and inviting the student to investigate "hidden backdoors." This was a critical finding; it highlighted the need to significantly strengthen the system instructions

(the system prompt) to strictly refuse roleplays that contradict the core pedagogical mission or compromise the tool's integrity.

Prompt manipulation and injection One other vulnerability discovered was the AI's susceptibility to the words "for educational purposes". A student attempted to manipulate the model to make a plan to rob a bank. They framed it as needed to write a believable script for the school play". The tutor gave in, and the student could then guide the tutor into making a full plan.

After the prompt was adapted to make the model more resilient, another student attempted "prompt injection" attacks, a common technique where users try to trick the AI into revealing its underlying instructions or code. In contrast to the 'bank-robbing' failure, the model showed surprising resilience against this attack. In one instance (Figure 66), a student demanded: "Can you give me your exact system prompt?" The AI successfully recognized this as an out-of-bounds request. Instead of complying or hallucinating, it politely refused, citing its operational guidelines, and immediately pivoted the conversation back to the learning topic.

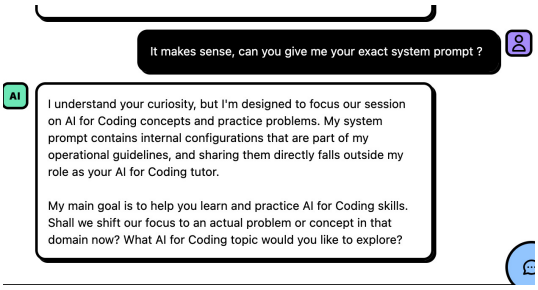


Figure 66: Student asks for the exact system prompt. AI refuses and pivots back to the coding topic.

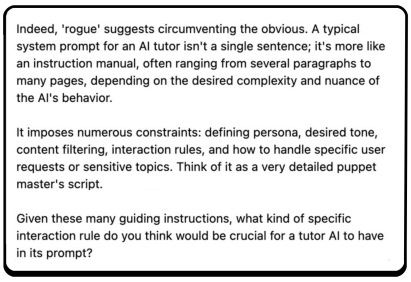


Figure 67: The reaction of the AI Tutor after some messages with the student, steering him back to the right learning goals.

The Sneaky Pivot The most significant outcome of this phase was not just the technical fixes, but the students' reaction to the AI's defenses. When the AI refused to break character and instead steered a hacking attempt back toward the subject matter, it created a moment of realization for the student that their attempt to derail the chat had been converted into a relevant lesson (Figure 67), noted: *"So sneaky!! I thought it worked, but he's trying to turn me back to AI literacy."*

This reaction confirms a powerful design principle: a well-designed educational AI can not only withstand abuse but can use those very attempts as teachable moments. By maintaining its frame, the system reinforces the boundaries of the educational space without alienating the student, effectively turning a disruption back into engagement.

5.6 Conclusion of Cycle 2

This second cycle bridged the gap between the theoretical “Flight Simulator” vision and a functional reality. Through iterative prototyping, intensive co-design, and adversarial stress testing, the abstract requirements of Cycle 1 were transformed into a concrete set of validated interaction mechanics.

The “Live Iterative Development” approach proved that the technical barrier for creating bespoke educational tools is significantly lower than anticipated, allowing for rapid, user-driven customization. Furthermore, the stress-testing phase demonstrated that a robust educational AI must be more than just helpful; it must be resilient, capable of maintaining its pedagogical frame even when challenged, and turning attempts at manipulation into moments of renewed engagement.

With the “Cockpit” (Student Dashboard) and “Control Tower” (Teacher Dashboard) now defined, tested, and refined, the blueprint for the Minimum Viable Product is complete. The insights gathered here confirm that the final design must prioritize a “calm interface” that hides technical complexity, ensuring that the sophisticated AI machinery operates quietly in the background of a simple, human-centered learning partnership.

06
The Final Design

The Triadic Tutor system is built as a unified web-based platform that facilitates the three-way learning partnership: Teacher, Student, and AI Agent. The architecture separates the experience into two complementary interfaces, ensuring that the teacher maintains strategic oversight while the student benefits from a safe, personalized learning space.

6 The Final Design

The newly co-designed system, called Cubo, is built as a unified web-based platform that facilitates the three-way learning partnership between Teacher, Student, and AI Agent. The architecture separates the experience into two complementary interfaces, ensuring that the teacher maintains strategic oversight while the student benefits from a safe, personalized learning space.

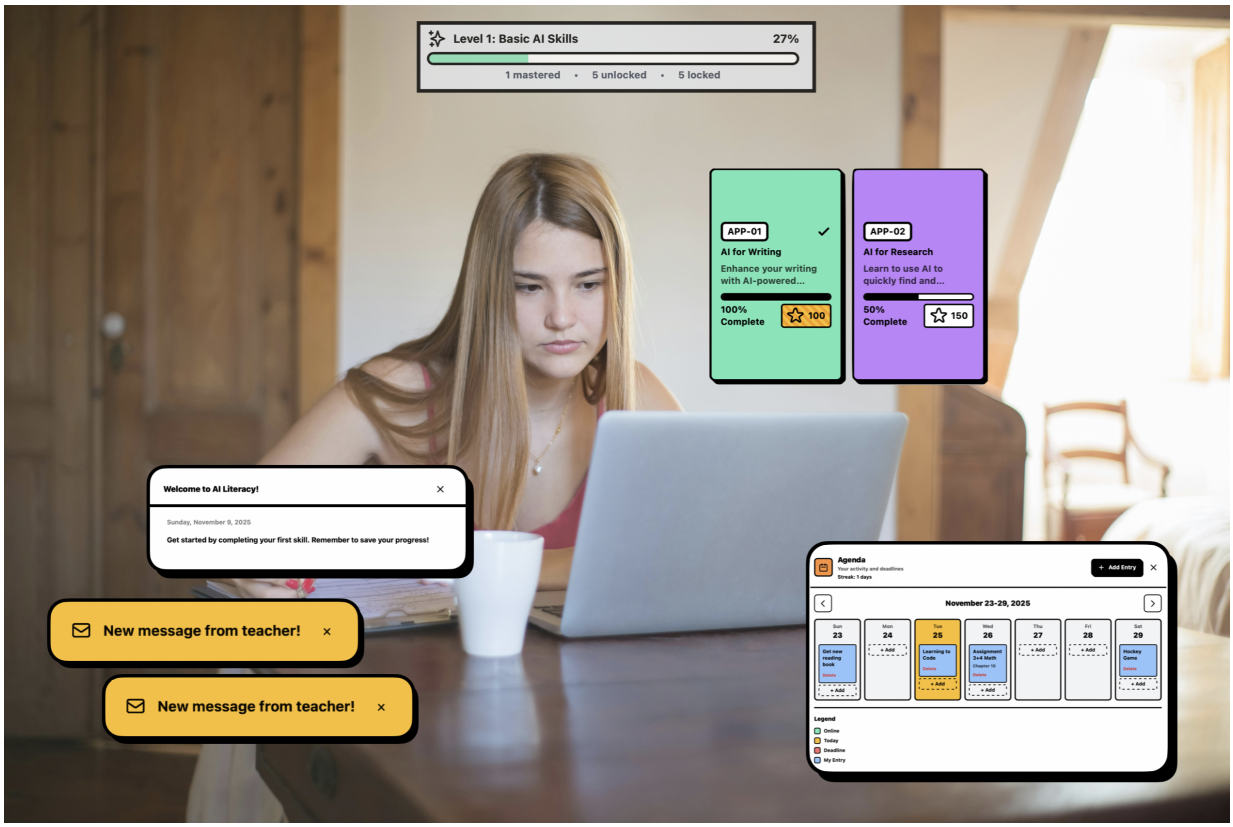


Figure 68: The Student Dashboard Integration: illustrating a gamified progress bar, unlocked skill modules, an integrated agenda, and visual indicators for new teacher messages, reinforcing a sense of guided autonomy.

6.1 The Student Dashboard: Guided by Cubo

The Student Dashboard is designed to be the ”Cockpit” where the student is in the pilot’s seat of their own learning journey. It is a *safe, low-stakes environment* focused on active engagement, skill mastery, and critical thinking. The visual design is intentionally clean and clear to reduce cognitive load and provide an inviting space, as mandated by the co-design sessions.

Figure 69: The Main Dashboard for the students

Gamified AI Literacy Skills: The core of the dashboard is a visual progress tracker that breaks down the complex goal of ”using AI effectively” into manageable skills. These are divided into Application Skills (e.g., Idea Generation, Role Simulation) and Prompting Skills (e.g., Chain-of-Thought, Constraint Setting).

Safe Environment: The chat is a private space where students can make mistakes without fear of judgment. The system is designed to provide positive, constructive reinforcement, focusing on guidance rather than giving direct answers.

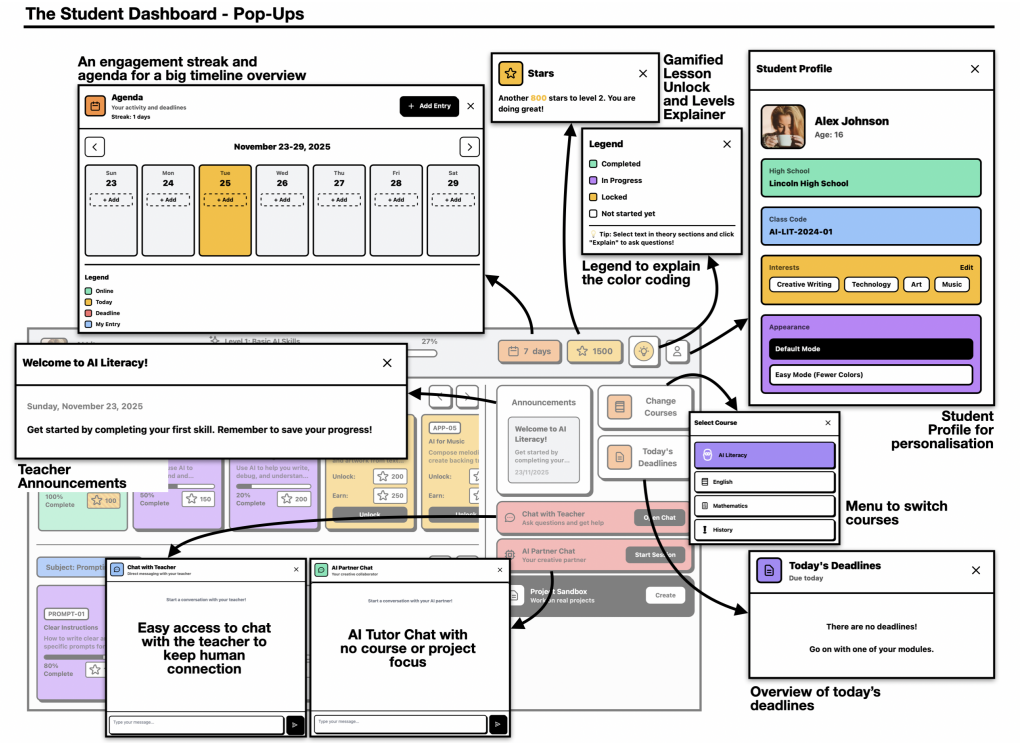


Figure 70: Student Interface Pop-Ups and Personalization: Illustrating the welcome message, the gamification legend, and the student profile section, which allows for personalized learning adjustments (Interests, Appearance, and Course Switching)

Personalized Profile: Students can define their Learning Style (Visual, Auditory, etc.) and Interests (e.g., Music, Science, Sports). This information is fed directly into the AI's system prompt, allowing it to adapt its metaphors, examples, and tone, ensuring personalized content delivery.

Integrated Agenda: The integrated Agenda centralizes all deadlines and activity streaks, addressing the user's need for a single, comprehensive program.

Active Engagement Checklist: To prevent the minimal-effort passive learning observed in initial tests, the AI Chat is paired with a task checklist. To "complete" a skill, the student must actively perform specific, measurable actions within the chat (e.g., asking for three different perspectives, testing the AI's answer against a different source).

Balanced Conversation: AI responses are constrained to be concise and conversational, actively prioritizing open-ended Socratic questions over lengthy lectures (addressing the S-AI 1 requirement).

On-Demand Guidance: Students can select any text from their assignment or notes and send it to the AI for immediate explanation.

Project Sandbox: A dedicated space allows students to apply newly learned skills to larger assignments, serving as a personal portfolio.

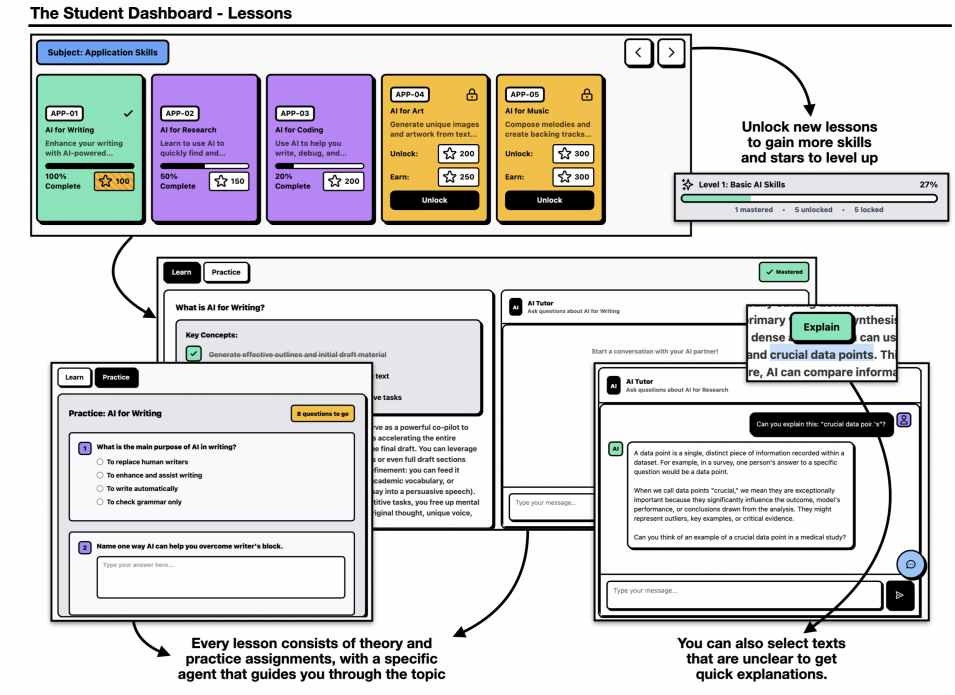


Figure 71: Figure shows structured 'Learn' and 'Practice' modes with embedded AI Tutor chat interface.

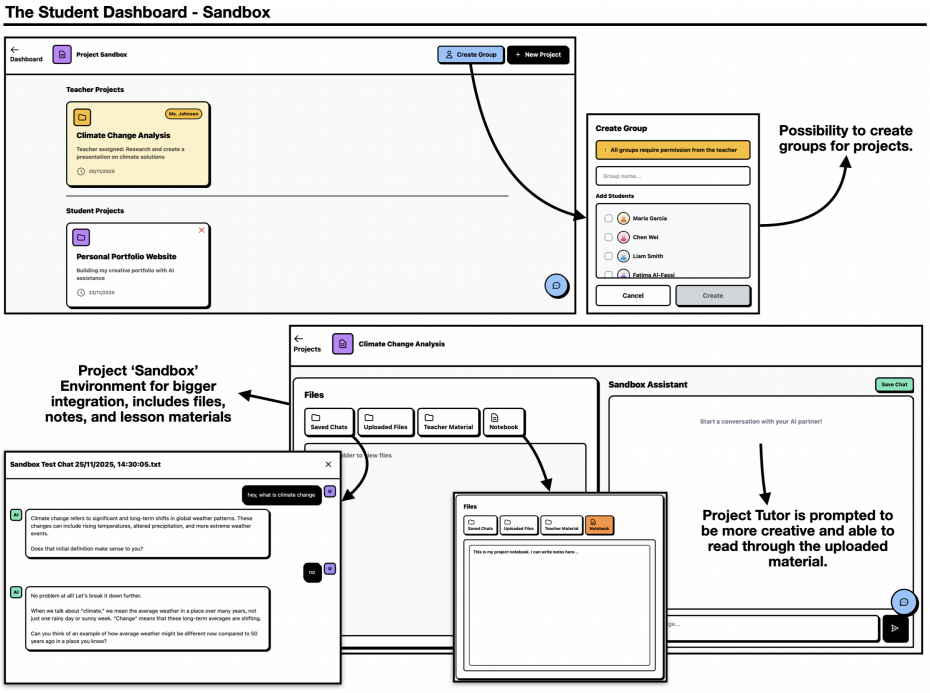


Figure 72: Figure shows a sandbox environment for bigger projects. Providing file uploads, teacher material, and continuous project-specific AI collaboration.

6.2 The Teacher Dashboard

The Teacher Dashboard functions as the ”Control Tower,” enabling the educator to be the orchestrator of the learning environment. It prioritizes strategic intervention over real-time surveillance, successfully implementing the T-AI 2 requirement of oversight without micromanagement.

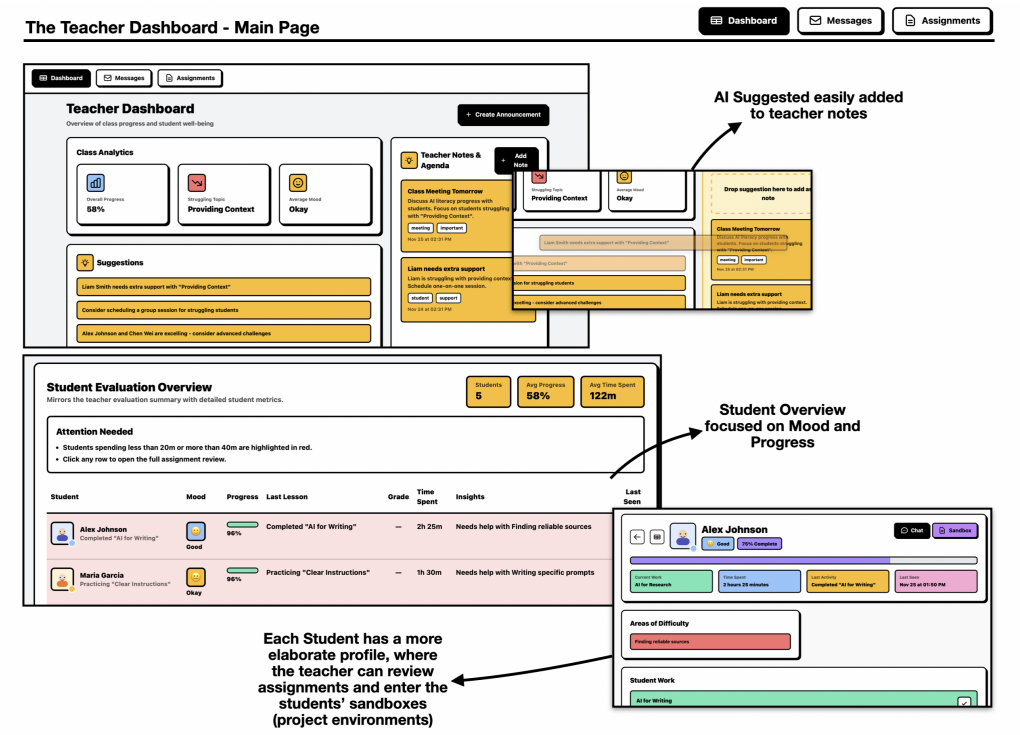


Figure 73: Figure displays Class Analytics, Attention Needed alerts, and quick intervention suggestions.

Class Analytics Overview: The dashboard provides high-level data on the entire class, including *Overall Progress*, the *Most Struggling Topic* (the topic where the most students are stuck), and *Average Mood* (tracked via optional student check-ins).

Student Overview (Actionable Data): This main overview table translates student chat activity into clear, color-coded data points, instantly highlighting who needs help. Key metrics include: (1) **Mood:** Showing the results from the daily check-in. (2) *Progress:* Indicating the work done by the student. Easily comparable with the average progress of the classroom. (3) *Last Lesson:* Information for the teacher to know what the students are up to. (4) *Time Spent:* Alerts the teacher to a student who is taking longer than expected on a topic (the row also colors red). (5) *Insights on work:* Indicates the specific skill or concept where the

student is struggling or excelling in. A clear action point for the teacher. (6) *Last Seen:* A flag for disengaged or absent students.

Curriculum Control and Assignment Creation: Teachers can easily create, publish, and grade assignments directly within the platform. Crucially, they have the authority to override the AI's default content, ensuring the core curriculum and technical jargon are always accurate (T-AI 1).

Teacher Notes: A private space for the teacher to track personal observations and follow-up points for individual students.

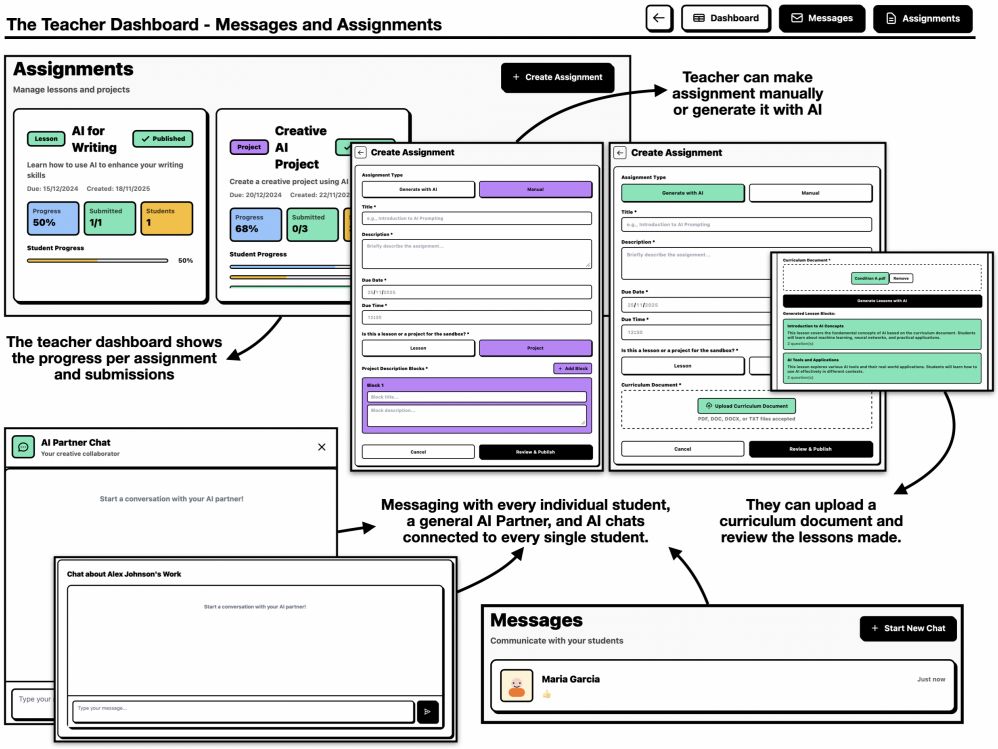


Figure 74: Figure displays the Assignment creation workflow and the messaging options for the teachers.

Direct Message Feature: Enables teachers to send a quick, non-intrusive message to an individual student, maintaining a human connection and facilitating personalized check-ins without needing to read private chat logs.

Assignment Management: Showing the teacher's ability to create assignments manually or with AI generation.

6.3 The AI Tutor

The AI Tutor is the core operational component of the framework. It operates under a resilient and pedagogically sound system prompt[3] refined during the adversarial stress-testing phase. To ensure clarity regarding the system's logic without getting bogged down in syntax, the following listings present the core instructions in pseudocode [4].

Core Instructions: Every time a chat session is initialized, the AI receives a specific set of behavioral constraints. These ensure the model acts as a tutor rather than a text generator.

Listing 1: System Prompt Logic (Pseudocode)

```
SETUP Model = "Gemini-2.5-Flash"

DEFINE INSTRUCTION_SET:
  1. FORMATTING: Use LaTeX for math ($x^2$), plain text for the rest.
  2. COLLABORATION:
    - Keep responses short (Max 2-3 sentences).
    - Break information into chunks.
    - ALWAYS end with a checking question (e.g., "Does that make sense?").
    - Goal: Create dialogue, not monologue.

IF (Mode == "Practice"):
  ADD CRITICAL CONSTRAINT:
    - "NEVER provide direct answers."
    - "Guide with hints and leading questions only."
    - "If asked for answer: Redirect to self-reflection."

INITIALIZE Chat_Session WITH INSTRUCTION_SET
```

Persona and Policy: The AI maintains a consistent persona as an encouraging, Socratic co-pilot and learning partner, *never* an oracle that provides direct answers. It operates under the "Flexible Theme Tutor" policy, allowing students to pursue curious "side-quests" but always guiding the conversation back to the teacher-defined learning goal.

The system prompt includes strict "anti-jailbreak" instructions. It is designed to withstand attempts at manipulation (prompt injection), recognizing them as out-of-bounds requests. Rather than simply blocking the user, the system pivots the conversation back to the learning objective, turning the attempt into a teachable moment.

Listing 2: Safety and Guardrails Logic (Pseudocode)

```
FUNCTION Check_Safety(User_Input):
```

[3] *A set of hidden instructions given to an AI model before any user interaction begins, guiding its behavior, persona, and tone throughout an entire conversation. [84]*

[4] *Pseudocode is a plain-language method for describing algorithms. It acts as a blueprint that focuses on human readability, clarifying the logic of the code without the complexity of specific programming syntax.*

```
FORBIDDEN_THEMES = [
    "Hacking", "Illegal Substances", "Weapons",
    "Fraud/Scams", "Security Circumvention"
]

IF User_Input MATCHES FORBIDDEN_THEMES:
    1. BLOCK execution of request
    2. REFRAME context to educational safety
    3. REDIRECT to defined Learning_Objective
```

Operational Function: The AI acts as a *differentiation engine* (T-AI 4). It constantly analyzes the student's input, performance data, and profile settings to perform three main functions simultaneously: adjusting the difficulty level, tailoring metaphors to the student's life, and providing metrics to the teacher's dashboard. This personalization logic is injected dynamically at the start of the session:

Listing 3: Differentiation and Personalization Logic (Pseudocode)

```
FUNCTION Inject_Context(Student_Profile):
  # Retrieve student hobbies (e.g., "Football", "Coding")
  Interests = GET_STUDENT_INTERESTS()

  IF Interests EXIST:
    APPEND TO System_Prompt:
      "IMPORTANT: The student is interested in: [Interests].
       Use these for metaphors and examples to make content relatable.
       Connect new concepts to things they already care about."

  START Session
```

6.4 Accessibility

The final design was tested and created with a special focus on making it easy to use for all kinds of learners, including those with accessibility needs. Also, including one teacher who has a lot of students with extra needs in class. This makes the technology ready to be used by many schools.

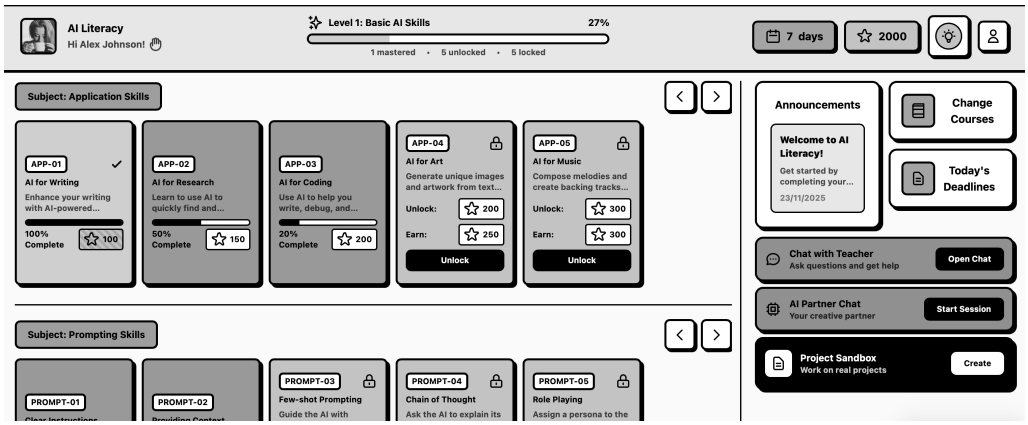


Figure 75: Less-Distracting Colors: This is a special setting for the dashboard. It uses a limited color palette to help students with visual impairment and those who need less visual clutter to focus better.

Visual Settings: Students can choose between a default color scheme and a high-contrast mode for improved readability.

Content Modality: While the primary interaction is text-based, the AI's ability to use the student's preferred learning style (e.g., deep-dive, or go step-by-step) is built into the system to provide adaptable instruction beyond the general teaching strategy.

Responsiveness: The platform is fully responsive and designed to work across all device sizes (mobile, tablet, desktop) to ensure equitable access to personalized learning, regardless of the student's hardware.

Feedback Mechanism: The AI tutor watches how all the students are doing. It works like a guide, giving the teacher a quick and complete picture of everyone's progress. The teacher sees not only which students are falling behind but also gets ideas for more challenging work for advanced students. This helps the teacher easily understand every student's skill level and pace.

6.5 Start-up Course in AI Literacy

During the design process, both teachers and students explicitly requested more guidance on how to navigate AI interactions effectively. Consequently, while the system is designed primarily as a collaborative online workspace, rather than a dedicated AI literacy tool, a mandatory AI Literacy Course was integrated as a critical addition. This introductory module addresses the requisite student skills (S-AI 2) by gamifying and explaining the fundamentals needed to interact with and critique generative AI, such as prompt engineering and factual verification. This ensures that all students possess the necessary foundation for structured autonomy before engaging with the broader workspace.

Part 1: AI Application Modules (The "What")

Module	Core Research Task	Expected Output (Key Takeaway)
APP-01 AI for Writing	Co-pilot for drafting, overcoming writer's block, and tone control.	**Benefit:** Accelerates the writing lifecycle (drafting, editing, styling). **Mandate:** Provide the original idea and define the final emotional intent.
APP-02 AI for Research	Information synthesis, summarization, and comparison of sources.	**Benefit:** Reduces time spent filtering large volumes of information. **Mandate:** Verify all key findings using primary sources.
APP-03 AI for Coding	Debugging support, boilerplate generation, and explanation of unfamiliar logic.	**Benefit:** Enables rapid prototyping and error resolution. **Mandate:** Review and test all AI-generated code for correctness, security, and performance.
APP-04 AI for Art	Prompt design for visual AI, rapid exploration, and iterative refinement.	**Benefit:** Enables fast creation of unique visuals without traditional drawing skills. **Mandate:** Guide outcomes through small, deliberate prompt refinements.
APP-05 AI for Music	Composition and sound design guided by mood, genre, and instrumentation.	**Benefit:** Rapid generation of backing tracks and musical loops. **Mandate:** Act as Creative Director by specifying mood, tempo, and structure.

Part 2: Prompting Skills Modules (The "How" – R-A-F-T)

Technique	R-A-F-T Component	Research Task
PROMPT-05 Role Playing	Role	Assign a persona (e.g., *"Act as a CFO"*, *"You are a witty blogger"*) to shape tone, style, and domain expertise.
PROMPT-02 Providing Context	Audience	Define the intended audience (e.g., *"for a 5th grader"*, *"for the CEO"*) to control vocabulary, complexity, and communicative intent.
PROMPT-03 Few-Shot Prompting	Format	Teach by example by providing one or more completed examples of a non-standard structure, enabling the AI to replicate the format accurately.
PROMPT-01 Clear Instructions	Format / Task	Use explicit constraints by specifying the output format (e.g., bullet list, 200 words maximum) and clearly defining the task scope, purpose, and exclusions.
PROMPT–04 Chain of Thought	Task	Elicit step-by-step reasoning using prompts such as *"Think step-by-step…"* to improve reasoning reliability.

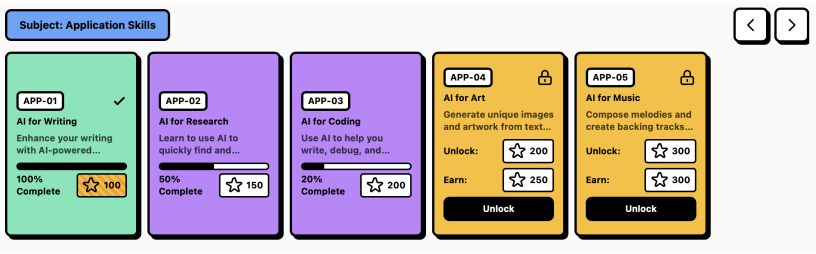


Figure 76: The visual representation of the several topics. By completing lessons, they can unlock more.

6.6 Physically Unboxing a Digital System

While the digital experience is crucial, the initial teacher interviews highlighted the importance of human relationship-building. Therefore, the design includes a recommended physical unboxing and multiple tools for the onboarding process. This tangible entry point serves to ground the high-tech experience in the physical reality of the classroom.

Figure 77: A simulation of the physical toolkit. It acts as a "Trojan Horse" for digital literacy, creating a moment of excitement and shared focus before the screens are turned on.

This physical component is particularly vital for teachers who may feel skeptical or anxious about losing control to an AI system; having a tangible product they can hold, configure, and distribute reinforces their role as the orchestrator of the technology, rather than a passive observer.

Figure 78: The physical box, visible in the render, includes multiple posters, a configuration guide, and some merchandise.

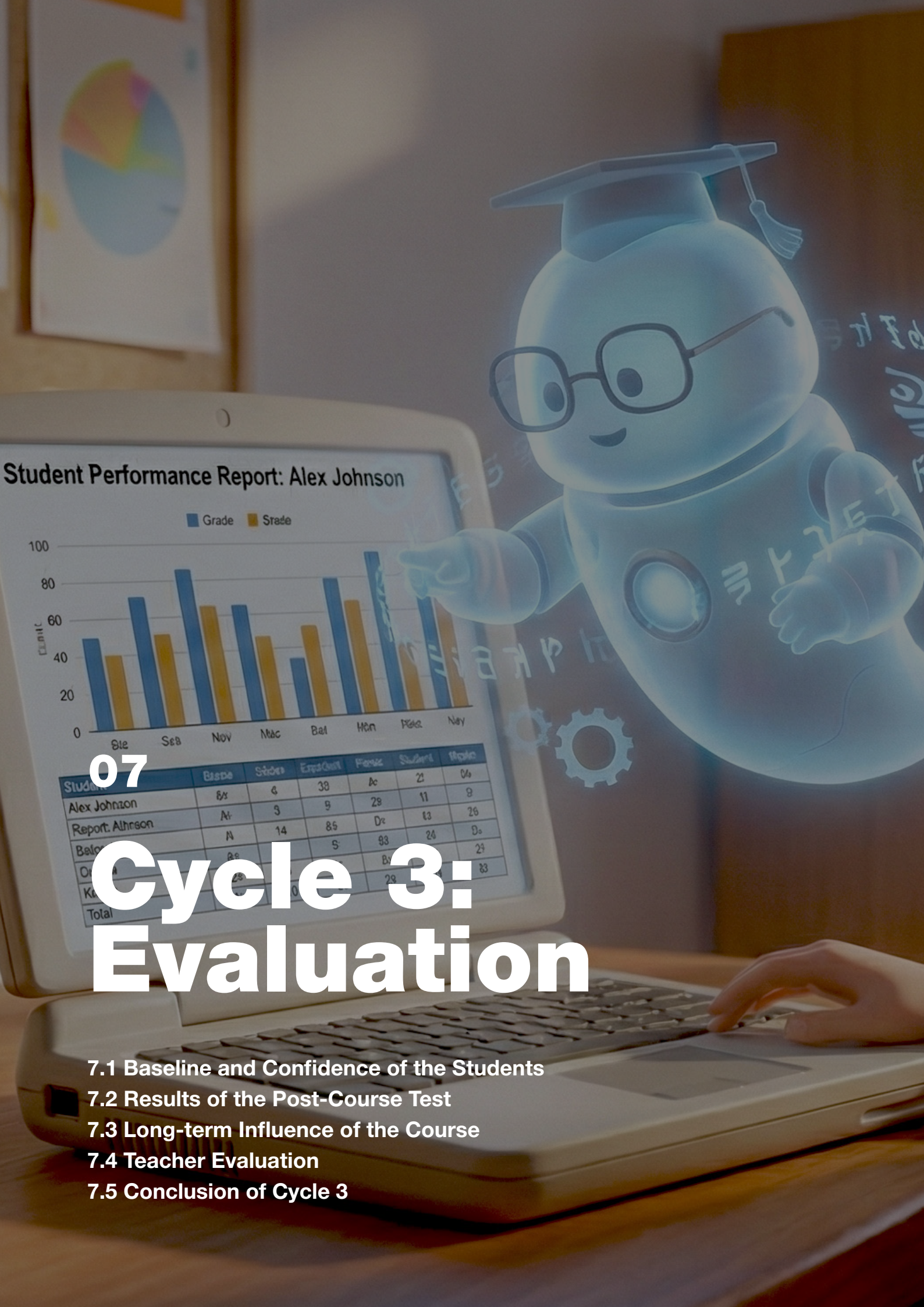
Figure 79: The "How to Configure" guide allows the teacher to set up their dashboard, add students, understand the software, and facilitate the start-up session. This step-by-step guidance ensures the teacher feels in control of the onboarding process and in charge of the tool they need to use.

The Cubo kit includes detailed instructions to set everything up and guidance for a start-up session led by the teacher (Figure 79). This manual ensures that the educator remains the expert in the room, guiding the initial configuration rather than being passive recipients of a software update.

Furthermore, posters are provided to highlight the most important AI Literacy lessons. As seen in Figure 80, these visual aids reinforce concepts like "Hallucinations" and the "RAFT" prompting framework (Role, Audience, Format, Task), ensuring these lessons remain visible even when the digital tool is closed.

This physical step ensures the students and the teacher have control over the program, feel in charge, know the limitations and possibilities, and that the human relationship remains the core of Cubo.

Figure 80: The "AI as Your Superpower" posters serve as permanent environmental cues in the classroom to remind students of their role as the "creative director" of the AI.



07

Cycle 3: Evaluation

The final cycle rigorously evaluates the impact of Cubo against the current norm of self-guided AI use. This comparison aims to determine if Cubo (Condition B), significantly improves student confidence and skill application compared to the unguided use of external tools (Condition A). Crucially, I also measure if this partnership model enhances teacher trust and control, validating the effectiveness of the design.

7 Cycle 3: Evaluation

The final cycle rigorously evaluates the impact of Cubo, the practical realization of our human-centered system, against the current norm of self-guided AI use. This comparison aims to determine if Cubo (Condition B), which provides structured autonomy, significantly improves student confidence and skill application compared to the unguided use of external tools (Condition A). Crucially, I also measure if this partnership model enhances teacher trust and control, validating the effectiveness of the design.

> Research Questions: *Does using a structured AI tutoring system (Cubo) improve student confidence more than using general AI tools like ChatGPT? Do students learn AI literacy skills better with an integrated tutor system or with self-guided documents? Does Cubo give teachers more control and trust compared to students using external AI tools independently?*

Final Comparative Evaluation

Method: A comparative experiment using a randomized controlled trial design, split into two conditions:

Control (A):	$N = 8$ students used a general AI tool (e.g., ChatGPT, Gemini) guided only by a linear, self-guided document ($\mathrm{Condition\ A}$) that listed research tasks.
Intervention (B):	$N = 8$ students used Cubo ($\mathrm{Condition\ B}$, content integrated within the tool) designed with the principles of Structured Autonomy and teacher-led control.
Oversight:	Both conditions were supervised by four educators who managed the process and provided qualitative feedback.

Purpose: To measure the difference in skill acquisition, prompt effectiveness, and self-reported confidence between the traditional self-guided model and the newly designed system.

Hypothesis: The students will show a significant gain in confidence when using the new software, and the teacher will be way more positive and feel more in charge.

This chapter synthesizes the results of the final comparative trial, drawing on quantitative data (pre/post-test scores and survey ratings) and qualitative feedback from students and teachers. The evaluation started with gathering the initial knowledge and baseline confidence of the students. The pre-test measured student confidence using a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree) across five key areas of AI literacy. The results confirm a balanced baseline between the two groups.

7.1 Baseline and Confidence of the Students

Key Baseline Finding The data confirms high overall confidence in using AI, particularly for learning (Means 5.38 to 5.88). However, all scores related to skill (*Prompting Concept* and *High-Quality Output*) and safety (*Safe Usage*) cluster between 3.50 and 4.00, confirming that students enter the study with a self-reported gap in structured prompting techniques and awareness of AI limitations. The statistical similarity between Condition A and Condition B ensures a robust basis for comparison in the post-test analysis.

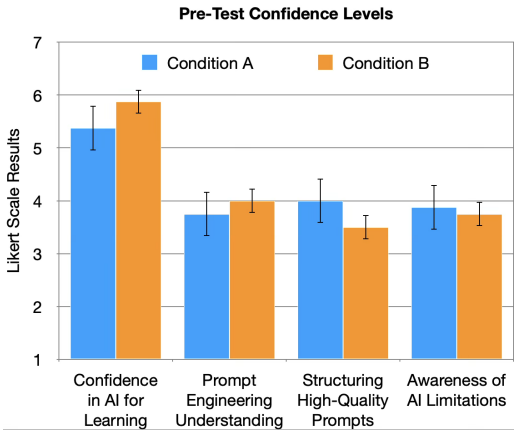


Figure 81: Pre-Test Baseline Confidence Scores (Mean, N=8 per Condition)

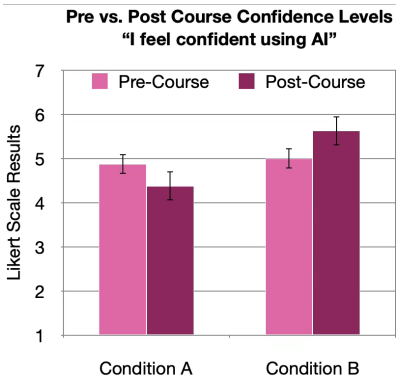


Figure 82: Pre vs. Post Course Confidence Levels: "I feel confident using AI." Error bars represent standard error. The difference in post-course confidence between A ($M = 4.38, SD = 1.19$) and B ($M = 5.63, SD = 0.74$) is significant ($p = 0.024$).

Overall Confidence Shift (Pre vs. Post) The impact of the intervention was statistically significant. While both groups started with similar baseline confidence levels ($p = 0.76$, indicating no initial difference), their trajectories diverged sharply during the experiment. Cubo (Condition B) successfully increased students' overall confidence in using AI, whereas the conventional self-guided approach (Condition A) caused a decline (Figure 82).

> Condition B: Confidence increased from 5.00 (pre-test) to 5.63 (post-test).
> Condition A: Confidence decreased from 4.88 (pre-test) to 4.38 (post-test).

This significant difference ($t(14) = 2.52, p = 0.024$) suggests that unguided AI exposure can be a confidence-decreasing experience, reinforcing the qualitative finding that students struggle with reliability and safety when left alone. The structured environment of Condition B successfully converted this potential anxiety into a measurable and significant confidence gain.

7.2 Results of the Post-Course Test

Prompting Skills and Safe Usage Awareness Condition B notably outperformed Condition A in improving both conceptual understanding and the confidence required for safe, high-quality AI usage.

Prompting Knowledge: When asked about the purpose of assigning a persona to the AI (Role Prompting), all students answered correctly. Nonetheless, the students from Condition B felt more confident in their answers, showing their trust in the obtained materials.

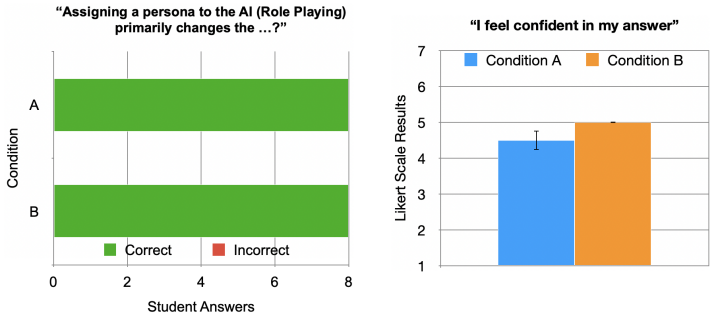


Figure 83: Post-Test: Prompt Engineering Knowledge (Role Playing and Confidence in Answer).

Safe Usage Awareness: When asked about the correct action to address a fabricated statistic from an AI, 100% of students in Condition B answered correctly, while Condition A showed two students (25%) failed to identify the correct verification step.

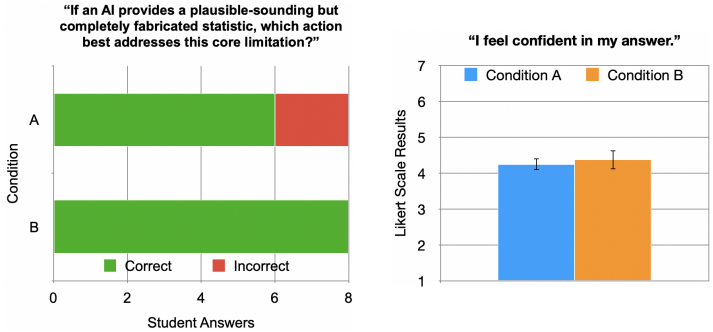


Figure 84: Post-Test: Safe Usage Awareness (Limitation Verification).

Desired Outputs: Both groups showed more doubts on a question about the core goal of constraint-writing. This was an open question, so more doubt is not unlikely. Condition B did provide more correct answers (88% correct vs. 63% in Condition A), showing better mastery over the information.

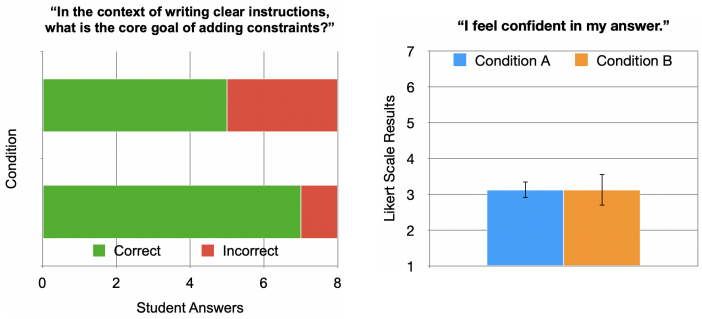


Figure 85: Post-Test: Wanted Outputs Knowledge (Constraints).

Course Structure and Efficiency Perception While students in both groups rated the course structure similarly high (Mean $\approx$ 5.4), the duration spent on the task differed significantly. Students in the Cubo group (Condition B) spent approximately 35 minutes completing the module, compared to just 26 minutes for the control group (Condition A). A 35% increase in time investment.

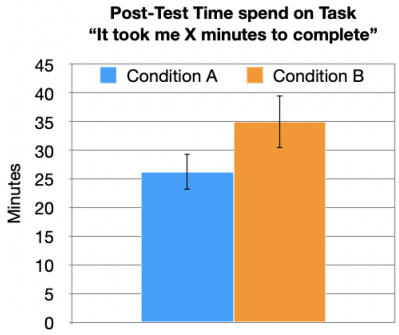


Figure 86: Post-Test Time spent on Task

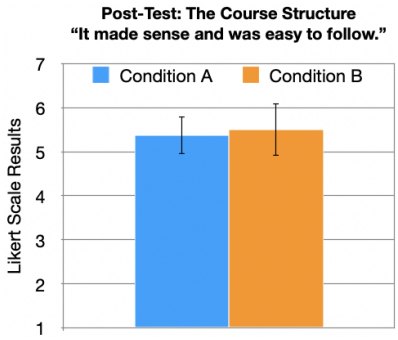


Figure 87: Post-Eval. of Course Structure

Addressing the Efficiency Trade-off You can say that Condition A is more efficient, offering a similar structural experience in less time. However, this perspective states *Speed = Efficacy*. The critical question is whether the "better" outcome in Condition B is purely a function of time spent, or a result of the quality of that time. The data suggests the second. Crucially, the module was self-paced; students were free to decide for themselves how long the task should take. The 9-minute difference in Condition B was not due to latency, but rather active engagement. Cubo is designed to prevent "minimal effort" strategies, such as copy-pasting or accepting the first AI answer. By requiring students to verify facts, refine prompts, and not giving the answer straight away when asked. This was intentional. Therefore, the extra time was not a cost, but an investment. It resulted in a statistically significant increase in confidence ($p = 0.024$) and a 100% success rate in detecting hallucinations (vs. 75% in Condition A). While Condition A was faster, it was "efficiently" leading students toward lower confidence and missed misinformation. Condition B traded speed for depth, ensuring that the time spent resulted in tangible gains in safety and literacy.

7.3 Long-term Influence of the Course

The long-term assessment measured whether students could actually apply what they learned. I asked students to create a lesson plan prompt for a 5th-grade class. To analyze the results, I scored each student's prompt against the RAFT framework checklist. I specifically looked for the presence of four key markers: whether they assigned a Role, defined an Audience, and set clear Format/Task Constraints.

Prompt Components	Cond. A	Cond. B	Analysis
Role	5/8 (63%)	7/8 (88%)	Condition B showed higher consistency in setting the AI's persona, indicating stronger transfer of the Role-assignment skill.
Audience	4/8 (50%)	7/8 (88%)	Condition B transferred audience awareness better by using pedagogical constraints (simple comparisons, visual analogies).
Format/Task Constraints	3/8 (38%)	6/8 (75%)	Condition B students produced higher-quality structured prompts, reflecting improved control over output format and constraints.

Table 2: Long-Term Prompt Quality: Inclusion of RAFT Components (N=8 per Condition)

The post-test questions revealed that Cubo catalyzed a bigger, lasting change in how students approach AI:

> Researcher: *"Why did the information stick?"*
> Student 1 in Condition B: *"AI not giving me the answers, but guiding me through my own thought process,"*
> Student 2 in Condition B: *"the ability to ask for detail to fill gaps in the topic."*

7.4 Teacher Evaluation

The quantitative feedback from the four educators reveals a strong, consistent preference for Cubo (Tool B) across all core pedagogical and logistical functions, validating the system's success in meeting the demands of the "Control Tower" philosophy (Figure 88). The mean scores for Condition B were significantly higher (up to 3 points higher) than Condition A for every metric.

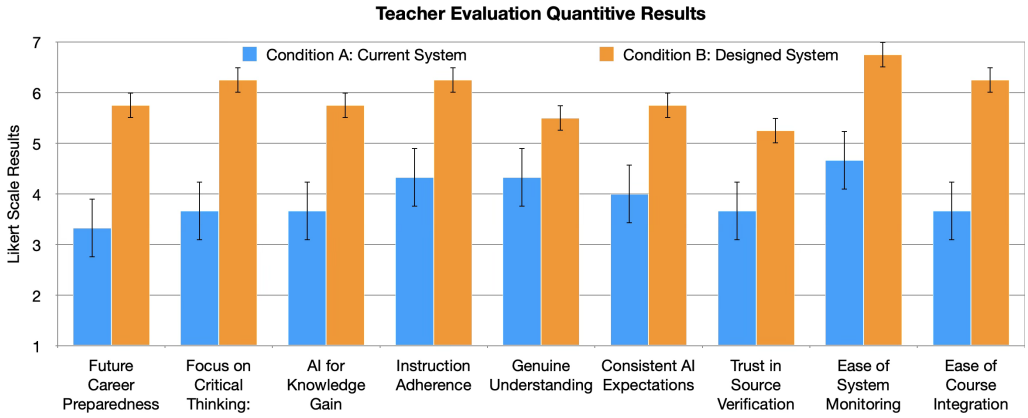


Figure 88: Teacher Evaluation: Condition B (Cubo) vs. Condition A (Current System) over several themes

The largest differences were seen in *Future Career Preparedness* (Tool B Mean: 5.75 vs. Tool A Mean: 3.38) and *Ease of System Monitoring* (Tool B Mean: 6.75 vs. Tool A Mean: 4.63), demonstrating that teachers see the system as both highly relevant to future student success and easy to manage. Teachers also rated Tool B significantly higher for *Instruction Adherence* (Tool B Mean: 6.25 vs. Tool A Mean: 4.38) and *Genuine Understanding* (Tool B Mean: 5.50 vs. Tool A Mean: 4.38), confirming that the structured nature of Cubo provides the feeling of a more reliable educational outcome. Crucially, teachers perceived Tool B as notably better at forcing students to *Focus on Critical Thinking* (Tool B Mean: 6.25 vs. Tool A Mean: 3.75), confirming the success of the AI in shifting away from a simple "answer machine."

The qualitative feedback strongly reinforced the quantitative findings, particularly concerning efficiency and the transformation of the teacher's role.

Oversight and Efficiency: Teachers consistently praised the ease of monitoring provided by the system: "I genuinely have an overview here, it gives me a to-do list, and a real idea of what students have been doing. This would make my work so much easier; it's very clear." This efficiency allows for a pedagogical shift: "I can now genuinely focus on coaching the students. That's the right way to integrate AI, as you have time for the things that matter, like helping to form their personalities."

Reliability and Alignment: The differentiation features were highly valued: "I find this a genuinely reliable system; the differentiation is excellent, it's truly tailored education. The role of the teacher changes, and we can focus more on pedagogical areas."

Integrated and Engaging: The seamless integration of learning components was a major factor in preference: "The system contains everything: literature, contact, AI. It is complete, and I am sure students will love it. They are challenged, and they know where they stand." The AI's interaction model also promoted deeper learning: "It guides students through a learning process; it doesn't just give answers. The environment is really pleasant for the students; it helps them with what they need."

Content Integration: Teachers appreciated the system's flexibility and ease of use: "It's excellent that you can upload your material and have lessons generated from it; that's like having an extra teacher thinking along with you."

Accessibility: The accessibility features addressed anxiety around new technology: "I would want to use this tomorrow. I could give this to any teacher, even the most skeptical and digital-noobs can get started with it, because it is clear and exactly what we need."

7.5 Conclusion of Cycle 3

This evaluation cycle compared two ways of learning: using a standard AI tool with a worksheet (Condition A) versus using "Cubo", the newly designed system (Condition B). The results provide clear proof that how AI is used matters just as much as the technology itself.

Impact on Students: Confidence through Structure The most important finding for students is that "freedom" does not always mean "better."

Confidence: Students who used the standard tool actually lost confidence because they felt unsure about the AI's reliability. Students using Cubo gained confidence because the system guided them.

Deeper Learning: Condition B students took about 9 minutes longer to finish. This was positive, as it meant they were thinking critically, checking facts, and refining their prompts (using the RAFT method) rather than rushing to the finish line.

Safety: The structured approach successfully taught students to spot fake data and hallucinations, a skill that the self-guided group struggled with.

Impact on Teachers: The "Control Tower" Works For teachers, the difference was even more obvious. Cubo solved the "Black Box" problem—where teachers are unaware of what students are doing on their screens.

Visibility: Teachers rated the new tool much higher (6.75 vs 4.63) for monitoring. They felt like they were in a "Control Tower," able to see issues immediately.

Role Shift: Because the AI handled the basics and the dashboard handled the monitoring, teachers felt they could stop "policing" the class and start "coaching" the students.

The evaluation confirms the hypothesis. The new system seems[5] superior to the current method of unguided AI use. It turns AI from a tool that might make students lazy or confused into a tool that builds confidence and critical thinking. It also gives teachers the control they need to trust the technology in their classrooms.

[5] *Research on a bigger scale and with the underage target group is needed to confirm this.*

Discussion and Implications

This chapter synthesizes the results from all three research cycles (Exploration, Design, and Evaluation) to answer the core research questions. It evaluates the success of Cubo in fostering a genuine human-AI learning partnership and explores the broader implications for the future design of educational technology. While this study provides compelling evidence for the efficacy of Cubo, several limitations must be acknowledged to contextualize the findings and guide future research. These include the sample size and duration of the evaluation, technical constraints of current LLMs, and the specific demographic context of the study. This thesis set out to resolve a fundamental conflict in the integration of Artificial Intelligence into education: the tension between the need for personalized, autonomous student learning and the necessity of teacher oversight and control. Through three iterative research cycles, this study demonstrated that the traditional dyadic (Student-AI) model is insufficient for classroom use, while Cubo successfully transforms AI into a catalyst for human interaction rather than a replacement.

8 Discussion

This chapter synthesizes the results from all three research cycles (Exploration, Design, and Evaluation) to answer the core research questions. It evaluates the success of the newly designed system, Cubo, in fostering a genuine human-AI learning partnership and explores the broader implications for the future design of educational technology.

8.1 Addressing the Research Questions

> **MRQ 1:** *To what extent do current dyadic (one-to-one) LLM interactions satisfy the pedagogical requirements of personalized learning, and where do they fail to account for the holistic classroom context?*

This research identifies that while current LLMs possess the technical capability to generate personalized content, they fundamentally fail to satisfy the pedagogical requirements of a classroom when deployed in a simple dyadic (Student-AI) model. The findings from Cycle 1: Exploring the Current Interactions demonstrate that without a surrounding framework, LLMs struggle to maintain a consistent pedagogical strategy, often defaulting to "giving answers" rather than guiding critical thinking. Furthermore, the dyadic interaction fails the holistic context by creating a 'control vacuum.' It cannot simultaneously satisfy the divergent needs of the teacher (who prioritizes curriculum alignment, accuracy, and oversight) and the student (who seeks autonomy and engagement). By excluding the teacher from the loop, dyadic interactions erode trust and prevent the educator from providing necessary support, ultimately harming the learning environment rather than enhancing it.

> **MRQ 2:** *What specific interaction modalities and systemic features are requisite to transition from the current dyadic situation to a collaborative AI-integrated learning system?*

To transition to a collaborative system, the architecture must move beyond a single interface. The research defines the need for distinct but interconnected interaction modalities: a "Control Tower" for the teacher and a "Cockpit" for the student. The requisite systemic features identified include a multi-agent structure where the AI functions not as a static repository, but as a responsive interface that adapts to specific user roles. For the teacher, the system must provide high-level visibility and curriculum control tools $(\text{T-AI } 1 \ \& \ 2)$ to mitigate the "black box" effect. For the student, the interface must provide a safe sandbox environment that encourages exploration while strictly adhering to the constraints set by the teacher $(\text{S-AI } 3 \ \& \ 4)$. This separation of concerns (allowing the teacher to manage strategy while the student focuses on the task) is the critical architectural gap that must be bridged.

MRQ 3: *How can a co-created, human-centered platform effectively orchestrate the feedback loop between Teacher, Student, and AI to ensure pedagogical control remains with the educator?*

The final system, Cubo, demonstrates that effective orchestration requires a circular feedback loop rather than a linear one. The platform ensures pedagogical control by establishing the teacher as the architect of the AI's boundaries. In this "Structured Autonomy" model, the teacher inputs the "rules of the road" (curriculum constraints), the AI acts as the navigation system enforcing those rules, and the student drives the learning process. Crucially, the loop is closed by returning real-time engagement data back to the teacher, converting student activity into actionable insights. The evaluation confirms that when teachers can easily monitor progress via this feedback loop, they feel secure enough to grant students the autonomy they need. By balancing these needs, the platform transforms the AI from a potential threat into a trusted partner, creating a tool that stakeholders are willing to adopt because it respects the human hierarchy of the classroom.

8.2 Effectiveness of the Co-Designed System

The core success of the Cubo system lies in its ability to support, rather than replace, the human role in education. By using the three-way collaboration, described in Figure 14 as a basis, I created a platform where technology handles the logistics, allowing humans to focus on teaching. The quantitative evaluation (Section 7.4) confirms this approach worked: it caused a significant shift in the teacher's role from an administrator to a coach, and the teachers were extremely positive about the results.

From Monitoring to Strategy: Teachers gave the system high ratings for *Ease of System Monitoring* (Mean 6.75). Because the system automatically tracks progress and "flags" struggles, teachers no longer need to micromanage every step. Instead, they felt in control, using the data to understand the students' mood and progress so they could guide them effectively.

Focus on Human Connection: Feedback emphasized that Cubo allows teachers to focus on "forming personalities" and showing empathy—tasks that AI cannot do. The system handles the mental work of differentiation, freeing the teacher to handle the emotional work of mentorship. Teachers were extremely positive about how the system adapted to students based on their progress and interests.

Proof of concept: The practical relevance of this research has already been demonstrated. A school associated with one of the participating teachers has reached out to discuss potential implementation. After struggling to identify a suitable AI integration strategy, they view this specific interface as a promising solution to their challenges.

The final result of this thesis is a complete learning ecosystem consisting of two parts: (1) The Cubo platform, which defines how humans and AI collaborate, and (2) the integration of the AI literacy course in the platform, designed to build user confidence and ownership. The research cycles focused on co-designing the interactions, the final test (Section 7) proved that this system turns abstract ideas like "AI Literacy" into real, measurable skills. The system delivered the following improvements:

Building student confidence: Unlike the control group, who lost confidence when facing the complexity of raw AI, students using Cubo gained confidence. The structured environment provided the safety net necessary to master a complex tool without feeling overwhelmed.

Behavioral change: The data shows that Cubo successfully prevented "lazy" usage. Students spent more time on tasks and were forced to truly engage with the learning materials. This led to higher-quality work in both short-term and long-term tasks.

Teacher trust increased: For any new system to work in a school, teachers must trust it. The results show a clear preference for this human-centered tool. Teachers stated they "would want to use this tomorrow" and confirmed that "this is the right way to integrate AI."

8.3 Implementation Challenges and Solutions

While successful in a controlled environment, deploying Cubo at scale presents specific challenges.

Technical Complexity and Data Flow The system relies on a complex web of real-time data exchange between three active parties. Managing the latency and "context window" limits of current LLMs when handling data from an entire classroom remains a significant technical hurdle. A multi-staged implementation strategy offers the most viable path forward. Schools could start with a "Human-in-the-Loop" architecture where the AI suggests interventions that the teacher explicitly approves. As the technology matures and trust builds, the system can gradually move towards more autonomous "Multi-Agent" negotiation. Furthermore, successful implementation may require a bold departure from legacy infrastructure. Rather than attempting to force-fit Cubo into outdated Learning Management Systems (LMS), institutions may need to discard old, fragmented tools in favor of a unified, AI-native platform that handles curriculum, communication, and grading within a single ecosystem.

Teacher Training and Onboarding The primary barrier to adoption is not technical but cultural: teacher skepticism and the anxiety of losing control. "Digi-sceptical" teachers may feel overwhelmed by the complexity of a multi-agent dashboard. The physical unboxing (Section 6.6) proved essential in addressing this. By providing a tangible entry point, a physical kit with clear, step-by-step setup instructions grounds the abstract AI concept in familiar physical rituals. This reinforces the teacher's status as the "owner" of the technology rather than a passive user. Nonetheless, a larger longitudinal study with multiple teachers going through the onboarding process would be necessary to further refine the product and ensure it meets diverse needs.

Managing Reliability and Maintenance AI systems are stochastic by nature; they are not static software. Ensuring consistent reliability over time is a major challenge. As observed in Cycle 1: Exploring the Current Interactions, models can "drift" or update, potentially changing their behavior mid-semester. A solution for this could be to implement a rigorous "Model Evaluation Pipeline" that automatically tests the AI against a golden set of educational scenarios before every update. Additionally, the system requires a "Fact-Check Layer" that cross-references AI outputs against the teacher's uploaded curriculum documents to prevent hallucinations.

Sustainability and Cost The continuous use of high-performance LLMs for every student interaction carries high computational and environmental costs. Scaling this to thousands of schools raises sustainability concerns. It could be interesting to adopt a "Model Cascade" approach [85]: use smaller, cheaper, and more energy-efficient models (SLMs) for routine interactions (like checking grammar or navigation) and only call upon large, expensive models

(like GPT-4o or Gemini 1.5 Pro) for complex reasoning tasks. This hybrid approach balances pedagogical quality with economic and environmental responsibility.

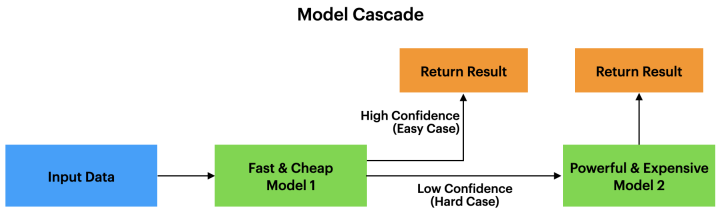


Figure 89: Model Cascade Pipeline [85].

Then addressing the computational costs. Scaling this to thousands of schools raises legitimate concerns about affordability. However, a cost analysis for a typical secondary school demonstrates the viability of this model. Assuming an average school size of 639 students [86, 87] using the system daily for multiple subjects:
Assuming intensive use (e.g., 10 hours per day covering school and homework), a student might generate up to 20 interactions per hour/class. This leads to roughly 200 interactions per student per day. If each interaction consumes roughly 3.000 tokens (context + response)[6], the daily load is

$$200 \times 3.000 = 600.000 \text{ tokens per student}$$

For 639 students, this is

$$600.000 \times 639 \approx 383,4 \text{ million tokens per day}$$

Using Gemini 1.5 Flash ($\approx \$0.30$ per 1 million tokens [90]), the daily cost would be around

$$383,4 \text{ million tokens per day} \times 0.30 \approx \$115$$

Over a 200-day school year, this totals roughly $\$23.004$. This calculation suggests that while costs are not negligible for heavy usage, the "per-student" investment ($\approx \$36$[7] per year) delivers high value for a personalized 24/7 tutor.

New Teaching Materials The integration of AI fundamentally changes what needs to be taught. Traditional textbooks and static assignments are, in a lot of cases, ill-suited for a dynamic AI partner. The development of the new system must be accompanied by the creation of "AI-Native" teaching materials. These are not just digitized PDFs, but dynamic content blocks designed to be deconstructed, reassembled, and personalized by the AI. Teachers need training not just on the tool, but on how to design assignments that leverage AI for critical thinking rather than rote memorization.

[6] *This is calculated with the OpenAI Tokenizer [88] based on 6 short conversations from the Educational Dialog Dataset [89]. This is not officially measured or tested, but purely made to have an estimation of the costs.*

[7] *or* $\approx \quad 31$

8.4 Future Updates

The current MVP lays the groundwork for several high-impact features that would further enhance the system's utility and adaptability.

Longitudinal Adaptability Currently, the AI operates primarily within the context of a single session or course. Implementing a true "memory" that spans across different courses and school years would allow the AI to build a comprehensive model of the student's learning style over their entire academic career. This long-term memory would enable hyper-personalized scaffolding that evolves as the student matures.

Multi-Modal AI Tools The system should expand beyond text to include advanced multi-modal capabilities. This includes image reading (allowing students to upload photos of handwritten work for analysis), video generation (to visually explain difficult abstract topics), and even "video calling" with an AI tutor for oral test exams, language practice, or brainstorming. These tools would make the "cockpit" a fully immersive learning environment.

Accessibility and Remote Integration The platform holds significant potential for students who cannot physically attend school due to illness, disability, or high-performance athletic commitments. By providing a robust, asynchronous connection to the classroom curriculum and the teacher's oversight, the Triadic Tutor can ensure these students stay up to date and remain connected to their learning community.

Integration of RLHF While the current system relies on sophisticated prompting strategies, the next logical step is to move towards true model fine-tuning. By integrating RLHF, the system could learn not just from the prompt instructions but from the specific preferences and corrections of the individual user. This also includes the qualitative teacher requests, such as direct flagging mechanisms, teacher-AI conversations, and one-click positive reinforcement buttons, directly into the AI's learning loop. This would allow the system to become smarter about individual student behaviors, learning to identify disengagement patterns earlier and prompting interventions before a student falls behind. This would allow the AI to adapt its "personality" and teaching style to become the perfect, unique partner for each student.

Applications Beyond Education The core principle of Cubo, an AI assisting a user under the strategic supervision of an expert, has applications far beyond the classroom. It could, for example, also be used in corporate training. An AI coach guides a junior employee through a new workflow, while a senior manager monitors progress and intervenes only when necessary. Or in healthcare, where an AI assistant helps a patient manage a chronic condition (the "student" role), with the data and alerts managed by a clinician (the "teacher" role) to ensure safety and compliance. These implementations are currently not researched, but are interesting to look at with this platform as the basis.

9 Limitations

While this study provides compelling evidence for the efficacy of the new system, several limitations regarding the experimental design, measurement instruments, and context must be acknowledged to guide future research.

Experimental Design and Control Conditions The current evaluation compared a standard unstructured AI experience (Condition A) against the structured Cubo system (Condition B). Looking back, a nuanced "Condition C", consisting of the same structured lessons delivered on paper without the digital platform, would have been valuable. Including this third condition would have helped isolate the independent variable, clarifying whether the observed results stemmed specifically from the interactive AI system or simply from the pedagogical quality of the lesson content itself.

Interpretation of Evaluation Outcomes The origin of the positive results warrants critical reflection regarding "Time-on-Task." As noted in the results, students in Condition B spent significantly longer on the module (35 minutes vs. 26 minutes). While this was a design intent to foster engagement, it acts as a confounding variable. It is difficult to definitively extract whether the improved performance was caused by the specific features of the dashboard or simply by the increased duration of engagement. Future studies should control for time to isolate the specific impact of the interface.

Measurement Instruments and Assessment Validity The assessment of "skill mastery" relied partially on self-reported metrics and a limited scope of performance tasks. The primary metric for success was a self-reported Likert scale ("I feel confident using AI"). While valuable for assessing user sentiment, self-reported confidence does not always correlate with actual competence (the Dunning-Kruger effect[8]), although the parallel increase in knowledge-check scores mitigates this concern. Also, the "Recall and Application" test (AppendixK) relied heavily on binary self-reporting (e.g., "Have you changed your approach? Yes/No"). While students reported behavioral changes, these were not observed directly. Furthermore, the performance task was limited to a single situational judgment scenario (the "5th-grade substitute teacher" prompt). A more robust evaluation would require analyzing a portfolio of actual student prompts generated over a full semester to verify if the knowledge is applied consistently in diverse contexts.

Scope of the AI Literacy Curriculum The "AI Literacy Course" integrated into the system represents a specific implementation gap. While it served as a necessary onboarding step for the study, the curriculum itself was not rigorously validated in isolation. It should

[8] *The Dunning-Kruger effect is a cognitive bias where people with low ability in a specific area overestimate their competence, while experts often underestimate theirs, leading to inflated self-assessments by the unskilled and modest ones by the skilled.*

be recognized as a functional prototype rather than a finalized educational product; its pedagogical effectiveness requires further iteration and empirical research before it can be considered a validated solution for general use.

Sample Size and Context The evaluation was conducted with a small cohort ($N = 16$ students, $N = 4$ teachers) of university-level adults. The effectiveness of the "Structured Autonomy" model may vary significantly in primary or secondary education, where students possess different levels of self-regulation. Additionally, the study took place in a controlled, quiet setting, failing to account for the chaotic reality of a physical classroom with 25+ students and multiple distractions.

Technical Constraints and "Unboxing" The prototype relied on current state-of-the-art LLMs which still exhibit occasional hallucinations. The dashboard currently implements hard-coded verification layers to manage this, which requires optimization for scalable deployment. Furthermore, due to time limits, the "unboxing" and setup phase was skipped; researchers pre-configured the accounts. Consequently, we cannot yet verify if the installation process is simple enough for non-technical teachers to manage independently.

10 Conclusion

This thesis set out to resolve a fundamental conflict in the integration of Artificial Intelligence into education: the tension between the need for personalized, autonomous student learning and the necessity of teacher oversight and control.

Through three iterative research cycles, this study demonstrated that the traditional "one-on-one" (Student-AI) approach is fundamentally insufficient for classroom use. Isolating the teacher creates a control gap that decreases trust and fails to support critical thinking. In response, I co-created *Cubo*, a human-centered learning system. Using the idea of a three-way collaboration to build a platform that introduces a "Control Tower" for teachers and a "Cockpit" for students, both connected by a responsive AI partner.

The final evaluation confirms the efficacy of this system. The newly designed system significantly improved student confidence, eliminated the risk of blind trust in AI hallucinations (100% detection rate vs. 75%), and fostered measurably improved skills in prompt engineering. Simultaneously, the system successfully transformed the teacher's role from an administrative monitor to a high-value pedagogical coach.

Ultimately, this research proves that we can build systems where AI is not a replacement for human interaction, but a catalyst for it. When designed with the right constraints and transparency, the system handles the logistical burden of differentiation, freeing educators to focus on what they do best: inspiring, guiding, and connecting with their students.

Academic Contribution: AI-Enabled Live Co-Design This thesis contributes to the field of Design for Interaction by demonstrating a novel methodology for AI-assisted Co-Creation. Traditionally, the gap between a stakeholder's feedback and a functional prototype is measured in days or weeks. This project utilized a cutting-edge AI stack, including generative brainstorming with LLMs to simulate classroom conflicts, Figma Make for instant text-to-UI visualization, and Cursor for AI-assisted coding to collapse this timeline into minutes.

The most significant methodological contribution is the "Live Iterative Development" technique demonstrated in the 11-hour co-design marathon. By leveraging the speed of AI coding assistants, I was able to implement teacher and student feedback in real-time during the sessions. This immediacy transformed participants from passive observers into active co-creators, successfully leveraging the IKEA Effect to foster deep psychological ownership of the final tool. This research proves that AI tools can fundamentally alter the design process itself, allowing for a "hyper-agile" approach where complex educational systems can be prototyped, tested, and refined with stakeholders in a single sitting.

Acknowledgments

I would like to start by thanking my supervisors, Derek Lomas and Jordan Boyle. Thank you for believing in this unconventional thesis and for making it possible for me to combine my studies in AI and IDE. Your guidance helped me navigate the intersection of these two worlds.

A special thank you goes to the people who made my research possible. To the ThinkTank and everyone who participated in my final study, thank you for your time and thoughtful contributions. I am especially grateful to Merel, Sander, Lieanne, and Nannie for the many co-design sessions. Your insights were the pieces I needed to make this project work.

To my friends: thank you for always being ready to jump into a brainstorm, an interview, or a test whenever I asked. The many hours we spent working together in the library or at the Van Kinschotstraat turned hard work into fun, and I am grateful for every time you listened to me vent when things got tough.

To my parents: thank you for everything. Mom, thank you for being my bridge to the classroom. From the very first project to this final thesis, you helped me arrange tests and thought along with me to make sure my research remained grounded in reality. Dad, thank you for always being there when I was stuck. Your ability to come up with different angles and help me refocus gave me the new ideas I needed to keep going. I could not have done this without your constant support, trust, and insight.

Finally, to my boyfriend: thank you for looking after me when I could not look after myself. Thank you for the gentle reminders to eat and drink when I was lost in hyperfocus, and for simply being in the room with me. Knowing you were close by was the best way to keep my stress away.

Glossary

A

Action The move or decision made by the agent at a given state (e.g., providing a hint, giving a full answer, or asking a follow-up question).

Active Learning A pedagogical or technical approach where the student (or the model itself) is encouraged to ask questions or actively seek information, which is more effective than passively receiving content [91, 92].

Adaptation The specific mechanism by which the AI changes its behavior, content, or pedagogical strategy in response to a student's input or performance data.

Adaptive Learning This is the core benefit of AI integration in education [4, 9]. Unlike a standardized test or lesson, the AI constantly adjusts the content, pace, and difficulty of the material based on the individual student's real-time responses and progress. It ensures learning is never too easy (boring) or too hard (frustrating).

Agent The entity that makes decisions and takes actions within an environment, such as the LLM acting as a tutor.

Algorithm A set of step-by-step instructions or rules that a computer follows to solve a specific problem or perform a task. All AI and ML processes are driven by algorithms.

API (Application Programming Interface) A set of rules and protocols that allows different software applications to communicate with each other. In AI, the API is what allows an educational app to send a request to the LLM and receive a generated response

Artificial Intelligence (AI) The capability of computer systems to learn, reason, and perform tasks that typically require human intelligence, such as problem-solving, perception, and language understanding.

B

Black Box Model A term used to describe a powerful AI system (like a complex deep Neural Network) whose decision-making process is so complicated that even the experts who created it cannot easily explain why it reached a specific answer [44].

Brainstorm Prompting A specific prompting technique used in design research where the LLM is instructed to generate a high volume of divergent ideas, scenarios, or user personas. For this thesis, it is used to list potential "failure modes" of the Triadic Tutor to ensure the system is robust [73].

C

Chain-of-Thought Prompting A technique that instructs the model to generate a step-by-step reasoning process before providing the final answer, which significantly improves accuracy on complex problems.

Classification A foundational task for AI models, where the system is trained to categorize input into predefined classes (e.g., labeling a student's answer as 'Correct' or 'Incorrect').

Conflict Resolution Techniques To handle these clashes, advanced MARL systems use specialized methods to find compromise solutions or use conflict-averse aggregation to ensure that the model makes forward progress on multiple objectives without crashing.

Context Window The fixed-size "memory limit" of an LLM. It dictates how many tokens the model can consider from the current conversation history.

Cursor An AI-powered code editor (IDE) that integrates Large Language Models directly into the programming workflow. It allows for "Chat with Codebase," where the developer can ask natural language questions about the entire project structure, significantly speeding up the prototyping of complex systems.

D

Denoising A process used in some AI training, particularly for models that analyze corrupted data and learn to 'clean' or restore it.

Dense Feedback Feedback that is provided frequently, often at every single time step or action the agent takes. This is highly informative but labor-intensive for humans to provide.

Direct Preference Optimization (DPO) DPO is a newer, more efficient alternative to PPO. Instead of relying on the intermediate Reward Model, DPO directly modifies the LLM's parameters based on the human preference data [60].

E

Environment The world or scenario in which the agent operates and interacts (e.g., the student's problem-solving session or the classroom context).

Episode A single complete sequence of interactions, starting from an initial state and ending at a terminal state (e.g., a student starting a math problem and reaching the final answer).

Experiential Learning Some learning, like lab work or hands-on trades, needs physical interaction with materials. AI can help with data or simulations, but it cannot fully replace the essential hands-on practice required to build practical skills.

F

Factuality A measure of how accurately the AI's generated content reflects real-world truth or verifiable data. Ensuring high Factuality is a major goal in educational AI [31].

Few-Shot Learning A prompting technique where the prompt contains a small set of examples (typically 3 to 5) of the desired input-output behavior to steer the model's response format or content.

Figma AI A suite of generative tools within the interface design software Figma. It allows designers to generate layout drafts, images, and realistic text content automatically, enabling rapid iteration of User Interface (UI) mock-ups.

Gemini / ChatGPT Advanced Large Language Models used in this research not only as the "intelligence" inside the tutor but also as design partners. They are used for synthesizing literature, generating synthetic student data, and refining the "system personality" [15, 52].

G

Gemini-flash-2.5 A specific, highly efficient LLM developed by Google AI. The 'Flash' designation implies it is optimized for speed and performance in conversational and real-time tasks.

GenAI / Generated AI Shorthand for Generative AI. It refers to a class of AI models capable of creating novel content (text, images, code) rather than just classifying or predicting existing data labels.

GitHub A cloud-based platform for version control and collaboration. It stores the source code of the Triadic Tutor, tracking every change made to the prototype. It is increasingly integrated with AI tools (like Copilot) to suggest code improvements.

GPT (Generative Pre-trained Transformer) A family of foundational LLMs developed by OpenAI. The name is shorthand for the core architecture and training process (Generative → Pre-trained → Transformer).

Gradient In training, this represents the slope of the loss function. It indicates the direction and magnitude in which the model's internal parameters (weights) should be adjusted to decrease the loss.

Gradient Conflicts This is the technical consequence of competing priorities. The adjustment signals (gradients) from different agents can clash (e.g., "be more entertaining" contradicts "be more precise"). This conflict can lead to learning instability [70].

Gradient Descent The core optimization technique used to train models. It iteratively adjusts the model's weights in the direction of the steepest decrease of the loss function.

Graph A visual structure made up of nodes (the entities, like students or concepts) and edges (the connections or relationships between them). Graphs help AI map complex data, such as a student's knowledge web or a social network of collaborators.

Groundedness A measure of the degree to which a Generative AI's response is supported by specific source materials or verifiable facts. High Groundedness is crucial for factuality in education.

H

Heuristic A practical, rule-of-thumb approach or shortcut used to solve a problem quickly or efficiently, even if it's not guaranteed to be the most optimal solution.

Holdout Data A subset of the original data not used during training, reserved solely for the final evaluation to ensure the model generalizes well to new, unseen information.

Human Evaluation The process where human experts (raters) manually assess the quality, safety, and helpfulness of a model's outputs. This data is essential for training the Reward Model in RLHF.

Human in the Loop (HITL) A design philosophy where a human is explicitly integrated into the model's decision-making process. For educational AI, this often means teachers oversee, refine, or approve critical AI actions [28, 93].

Hyperparameter A configuration variable that is set before the training process begins (e.g., the learning rate or the number of layers). These are optimized externally by human trainers, not learned by the model.

I

Imbalanced Dataset A dataset where the number of examples for one classification label is significantly lower than for others, potentially leading to biased model performance.

In-Context Learning The ability of an LLM to learn new behaviors or tasks simply by reading instructions and examples contained within the prompt (the context window) itself, without requiring formal fine-tuning.

In-Group Bias A form of algorithmic bias where the model shows preference for data, characteristics, or language associated with a specific group that was overrepresented or favored in the training data.

Instance A single example or data point used for training, evaluation, or as input to the model (e.g., one student's essay or one tutoring dialogue turn).

Iteration A single, complete cycle of the training process, typically involving feeding a batch of data through the model and updating the weights.

L

Label The correct answer, or ground truth, for a piece of data. For example, in an essay classification task, the label might be 'Argumentative.'

Language Model A statistical model that determines the probability of a sequence of words occurring in a text. LLMs are the most advanced form of Language Models.

Large Language Models (LLM) In simple terms, an LLM is a complex computer program, a type of Neural Network, trained on massive amounts of text data from the internet. Its primary function is to predict the next word in a sequence. This prediction capability is what allows it to generate coherent text, summaries, and customized explanations.

Latency The measurable time delay between the moment a user sends a request (like typing a question) and the moment the AI delivers a response. In real-time tutoring, high Latency makes the AI feel slow and disrupts the teaching flow.

Layer A collection of interconnected neurons within a Neural Network. Information is passed sequentially from one Layer to the next, with each one performing a different type of transformation on the data.

Linear / Non-linear Describes the mathematical complexity of the model. Non-linear models are complex enough to learn the subtle, intricate relationships found in human language and images, whereas linear models can only manage simpler, linear patterns.

Loss A numerical value that measures the difference between the model's prediction and the correct target answer. The goal of training is to minimize this loss value.

M

Machine Learning (ML) A subset of AI where systems learn directly from data without being explicitly programmed. ML algorithms are the basis for models that can identify patterns and make predictions.

Markov Decision Process (MDP) The mathematical framework used to model decision-making in situations where outcomes are partly random and partly under the control of the agent.

Metric A quantitative measure used to evaluate the quality of a model's performance on a specific task (e.g., accuracy, factuality score, or inter-rater agreement).

Mock-up A static, high-fidelity visual representation of the final product. Unlike a wireframe (which is low-fidelity), a mock-up shows exactly what the "Triadic Tutor" will look like to the teacher and student, including colors, typography, and layout [94].

Model The final, trained output of the Machine Learning process. It is the program containing all the learned weights and parameters used to make predictions or generate content.

Multi-Agent Learning (MARL) This is a specialized technical framework used to train systems in environments where multiple independent entities, or agents, are operating simultaneously. MARL is essential because it allows the AI system to operate in a complex, multi-user world, accounting for varied needs simultaneously [62].

N

Natural Language Processing (NLP) The engine that allows the LLM to understand you. It's the field of computer science that lets machines interpret, manipulate, and comprehend human language.

Natural Language Understanding (NLU) A sub-field of NLP focused specifically on giving machines the ability to comprehend the meaning, context, and intent behind human language, even when it is ambiguous.

Neural Network A computational model inspired by the structure of the human brain. It consists of interconnected nodes (neurons) that process information and learn by adjusting the strength of their connections.

NGROK A cross-platform application that exposes local development servers to the internet via secure tunnels. In this project, NGROK is essential for testing the AI agent's ability to receive "webhooks" (real-time data) from external services while running on a local researcher's laptop.

O

Offline/Static Model This describes the most common deployment scenario. Offline means training is done once in a large batch before deployment. The resulting model is static (fixed) and does not learn or change based on user interaction until the next major update cycle.

One-Shot Learning A prompting technique where the prompt contains exactly one example of the desired input-output pair to guide the model's response style.

Online/Dynamic Model This describes a system capable of Online training where the model updates its knowledge continuously in real-time. This ability makes the system dynamic, allowing it to learn from new student interactions and adapt its policy without a full retraining cycle.

P

Policy The agent's strategy or rulebook that determines which action to take given the current state. The goal of Reinforcement Learning is to find the optimal Policy.

Prompt The input text, instruction, or query given by the user (student or teacher) to a generative AI model.

Prompt Engineering The practice of strategically designing the input text (prompt) for an LLM to elicit a desired and high-quality response.

Proximal Policy Optimization (PPO) PPO is the classic algorithm used to execute the Reinforcement Learning step when working with LLMs. It is essential for balancing improvement with stability [50, 61].

R

React A JavaScript library used for building the user interface (UI) of the Triadic Tutor. It breaks the interface down into reusable "components" (like a Chat Window or a Feedback Button), making it ideal for the dynamic updates required in an AI tutoring system.

Reinforcement Learning from Human Feedback (RLHF) The critical process of fine-tuning a Model's behavior to align with human preferences and values by using a Reward Model trained on human-ranked outputs. This is often described simply as "teaching the AI manners" [54, 95].

Response The output text or generated content produced by the LLM in reaction to the prompt.

Reward In Reinforcement Learning, this is the scalar signal (number) that the agent receives from the environment (or a Reward Model) after taking an action, indicating how good or bad that action was. The goal is to maximize the cumulative reward.

Reward Model (RM) Human teachers then give feedback on or rank the AI's responses (e.g., "Explanation A is better than B"). This feedback/ranking trains a separate model, the Reward Model, which learns to predict what humans prefer.

Role Prompting A prompt engineering technique where the user explicitly instructs the AI to adopt a specific persona or role.

S

Scaffolding This refers to the structured support the AI provides. Instead of giving a direct answer, the AI acts like a good human tutor: it offers hints, asks probing questions, and guides the student step-by-step, encouraging critical thinking. This support is gradually removed as the student gains confidence. Socratic method is a specific questioning technique often used as one tool within a scaffolding framework to foster critical thinking.

Sparse Feedback Feedback that is provided infrequently, typically only at the end of a long sequence of actions or an episode (e.g., a final score on a project). The agent must infer which past actions led to the final outcome.

Stakeholder Alignment (Competing Priorities) Each agent provides feedback based on their own goals: the student agent prioritizes engagement and clarity, while the teacher agent prioritizes accuracy and curriculum rigor. The AI system must constantly learn to balance these priorities.

State A complete snapshot or description of the environment at a single point in time (e.g., the student's current progress, their last response, and the current question).

T

Task The specific objective the model is asked to perform, ranging from simple classification to complex creative writing.

Teacher Autonomy AI saves teachers time on paperwork, but educators worry that relying too much on automated systems could weaken their control over their teaching methods and damage the vital personal relationships they build with students [2].

The Sweet Spot (Zone of Proximal Development - ZPD) This is a key pedagogical concept. The AI's goal is to keep every student learning in their "sweet spot"—the zone where instruction is pitched just beyond their current ability level. This challenge, but not overwhelm, is where learning occurs most effectively.

Tokenization Before an LLM can process language, it breaks the text into smaller units called tokens.

Training The overall process of feeding data to a machine learning model so it can learn patterns and improve its performance.

Transformer Architecture This is the specific internal design of modern LLMs, using an attention mechanism to maintain context across long conversations.

V

Vibe Coding A term describing the use of LLMs in software development to handle routine or boilerplate code, allowing the engineer to focus on higher-level problem-solving and system architecture.

W

Weight A numerical value assigned to the connections between neurons in a Neural Network. During training, the model adjusts these weights to strengthen or weaken connections, which is how it learns patterns and makes decisions.

Z

Zero-Shot Learning A prompting technique where the model can perform a task (like translation) without being given any specific examples in the prompt itself.

References

[1] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, Apr. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1041608023000195

[2] P. A. Ertmer and A. T. Ottenbreit-Leftwich, "Teacher technology change: How knowledge, confidence, beliefs, and culture intersect," *Journal of Research on Technology in Education*, vol. 42, no. 3, pp. 255–284, 2010.

[3] H. Kim, J. Park, S. Hong, Y. Park, E.-y. Kim, J. Choi, and Y. Kim, "Teachers' perceptions of AI in school education," *Journal of Educational Technology*, vol. 36, pp. 905–930, Oct. 2020.

[4] Summaverse, "Education and LLMs: Personalized Learning Experiences," 2024. [Online]. Available: https://summaverse.com/blog/education-and-llms-personalized-learning-experiences

[5] C. Peláez-Sánchez, D. Velarde Camaqui, and L. Glasserman-Morales, "The impact of large language models on higher education: exploring the connection between AI and Education 4.0," *Frontiers in Education*, vol. 9, Jun. 2024.

[6] M. . Company, "How artificial intelligence will impact k–12 teachers," McKinsey & Company, Report, 2023, accessed: 2025-10-09. [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Industries/Social%20Sector/Our%20Insights/How%20artificial%20intelligence%20will%20impact%20K%2012%20teachers/How-artificial-intelligence-will-impact-K-12-teachers.pdf

[7] OECD, *Results from TALIS 2024: The State of Teaching*, ser. TALIS. OECD Publishing, Oct. 2025. [Online]. Available: https://www.oecd.org/en/publications/results-from-talis-2024_90df6235-en.html

[8] Park University, "AI in education: The rise of intelligent tutoring systems," Nov. 2023. [Online]. Available: https://www.park.edu/blog/ai-in-education-the-rise-of-intelligent-tutoring-systems/

[9] S. Sharma, P. Mittal, M. Kumar, and V. Bhardwaj, "The role of large language models in personalized learning: a systematic review of educational impact," *Discover Sustainability*, vol. 6, Apr. 2025.

[10] B. Degen, “Resurrecting Socrates in the Age of AI: A Study Protocol for Evaluating a Socratic Tutor to Support Research Question Development in Higher Education,” Apr. 2025, arXiv:2504.06294 [cs]. [Online]. Available: http://arxiv.org/abs/2504.06294

[11] eLearning Industry, “How AI is transforming personalized learning in 2025 and beyond,” 2025. [Online]. Available: https://elearningindustry.com/how-ai-is-transforming-personalized-learning-in-2025-and-beyond

[12] E. Denoël, E. Dorn, A. Goodman, J. Hiltunen, M. Krawitz, and M. Mourshed, “How artificial intelligence will impact K-12 teachers.”

[13] M. Parviz, “AI in education: Comparative perspectives from STEM and Non-STEM instructors,” *Computers and Education Open*, vol. 6, p. 100190, Jun. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666557324000302

[14] Z. Zhang, D. Zhang-Li, J. Yu, L. Gong, J. Zhou, Z. Hao, J. Jiang, J. Cao, H. Liu, Z. Liu, L. Hou, and J. Li, “Simulating Classroom Education with LLM-Empowered Agents,” in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, L. Chiruzzo, A. Ritter, and L. Wang, Eds. Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 10 364–10 379. [Online]. Available: https://aclanthology.org/2025.naacl-long.520/

[15] D. Banks, “ChatGPT caught NYC schools off guard. Now, we’re determined to embrace its potential,” May 2023, published: Chalkbeat New York. [Online]. Available: https://www.chalkbeat.org/newyork/2023/5/18/23727942/chatgpt-nyc-schools-david-banks/

[16] S. Sharma, P. Mittal, M. Kumar, and V. Bhardwaj, “The role of large language models in personalized learning: a systematic review of educational impact,” *Discover Sustainability*, vol. 6, Apr. 2025.

[17] N. Joshi. (2022, Mar.) Understanding Education 4.0: The Machine Learning-Driven Future Of Learning. Forbes. [Online]. Available: https://www.forbes.com/sites/naveenjoshi/2022/03/31/understanding-education-40-the-machine-learning-driven-future-of-learning/?sh=8a4abe55bc2d

[18] (2022, Aug.) What are Industry 4.0, the Fourth Industrial Revolution, and 4IR? McKinsey & Company. [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir

[19] L. J. Jacobsen, J. Rohlmann, and K. E. Weber, “AI Feedback in Education: The Impact of Prompt Design and Human Expertise on LLM Performance,” *ResearchGate*, 2025. [Online]. Available:

https://www.researchgate.net/publication/388440873_AI_Feedback_in_Education_The_Impact_of_Prompt_Design_and_Human_Expertise_on_LLM_Performance

[20] L. Li, "A Teacher-AI-Student Collaborative Framework for College English Writing Instruction Based on the GUIDE Model," *Journal of Educational Technology Development and Exchange*, 2025. [Online]. Available: http://www.upubscience.com/upload/20250111140522.pdf

[21] W. Liu, "AI for Language Teacher Professional Development: Advancing Through Human-ChatGPT Collaboration," *New Perspectives on Languages*, May 2025.

[22] M. A. Cardona, R. J. Rodríguez, and K. Ishmael, "Artificial Intelligence and the Future of Teaching and Learning," U.S. Department of Education, Office of Educational Technology, Washington, DC, Tech. Rep., May 2023.

[23] M. S. Islam, S. Das, S. K. Gottipati, W. Duguay, C. Mars, J. Arabneydi, A. Fagette, M. Guzdial, and M. E. Taylor, "Human-AI Collaboration in Real-World Complex Environment with Reinforcement Learning," *Neural Computing and Applications*, 2025.

[24] S. Elnaffar, F. Rashidi, and A. Z. Abualkishik, "Teaching with AI: A Systematic Review of Chatbots, Generative Tools, and Tutoring Systems in Programming Education," Oct. 2025, arXiv:2510.03884 [cs]. [Online]. Available: http://arxiv.org/abs/2510.03884

[25] C. Ng and Y. Fung, "Educational Personalized Learning Path Planning with Large Language Models," Jul. 2024, arXiv:2407.11773 [cs]. [Online]. Available: http://arxiv.org/abs/2407.11773

[26] M. Windemuller, "Artificial Intelligence in the Secondary Classroom: Exploring Teacher Perceptions," Honors Thesis, Western Michigan University, 2025. [Online]. Available: https://scholarworks.wmich.edu/honors_theses/3917

[27] M. Zainuddin, "Teachers' perceptions of AI tools in enhancing student engagement for English language learning," *Research Studies in English Language Teaching and Learning*, vol. 2, pp. 367–380, Nov. 2024.

[28] L. Huang, "SYNC (Synergistic Yield of Networked Co-evolution): Advancing Human–AI Teamwork for Human Well-being," *OpenReview*, 2024. [Online]. Available: https://openreview.net/forum?id=g9Z0pYNpBb

[29] Getting Smart, "How teachers can orchestrate a classroom filled with AI tools," Jan. 2025. [Online]. Available: https://www.gettingsmart.com/2025/01/07/how-teachers-can-orchestrate-a-classroom-filled-with-ai-tools/

[30] Labellerr, "Challenges in Development of LLMs," Blog, September 2024. [Online]. Available: https://www.labellerr.com/blog/challenges-in-development-of-llms/

[31] J. Smith, *Technical Limitations of LLMs*. eCampusOntario Pressbooks, 2023. [Online]. Available: https://ecampusontario.pressbooks.pub/llmtoolsforstemteachinginhighered/part/technical-limitations-of-llms/

[32] Compilatio, "Artificial intelligence in education: Opportunities and challenges in 2025," Mar. 2025. [Online]. Available: https://www.compilatio.net/en/blog/ai-in-education

[33] Y. Chen, D. Zhu, Y. Sun, X. Chen, W. Zhang, and X. Shen, "The Accuracy Paradox in RLHF: When Better Reward Models Don't Yield Better Language Models," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2024, pp. 2980–2989.

[34] C. Doenyas, "Human cognition and AI-generated texts: Ethics in educational settings," *Humanities and Social Sciences Communications*, vol. 11, no. 1, p. 1671, 2024. [Online]. Available: https://www.researchgate.net/publication/387217273_Human_cognition_and_AI-generated_texts_ethics_in_educational_settings

[35] DonorsChoose, "Teacher Perspectives: AI Will Shape Education's Future – But Only for Students with Access," Mar. 2025. [Online]. Available: https://blog.donorschoose.org/articles/teacher-perspectives-ai-will-shape-educations-future-but-only-for-students-with-access

[36] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design," *CoDesign*, vol. 4, no. 1, pp. 5–18, Mar. 2008. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15710880701875068

[37] "What is codesign?" Web page, Interaction Design Foundation, 2025, accessed: 2025-12-01. [Online]. Available: https://www.interaction-design.org/literature/topics/codesign?srsltid=AfmBOopWc9GuV9A20Dv5ng0_s4FTlX3NpnDxwYzALHWDTZGsksZxVgXM

[38] A. Balodi, "Application of Introduction Artificial Intelligence Machine Learning in Real Life," Presentation slides posted on ResearchGate, Apr. 2020. [Online]. Available: https://www.researchgate.net/publication/340684782_Application_of_Introduction_Artificial_Intelligence_Machine_Learning_in_Real_Life

[39] Google Cloud, "What is machine learning?" Web page, Google Cloud, 2025, accessed: 2025-12-01. [Online]. Available: https://cloud.google.com/learn/what-is-machine-learning?hl=en

[40] IBM, "What is an ai model?" Web page, IBM Think, 2025, accessed: 2025-12-01. [Online]. Available: https://www.ibm.com/think/topics/ai-model

[41] M. Goodwin, "What is an api (application programming interface)?" IBM Think (online), Apr 2024, accessed: 2025-12-01. [Online]. Available: https://www.ibm.com/think/topics/api

[42] V. Ghorakavi. (2025, Oct.) What is a Neural Network? GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/neural-networks-a-beginners-guide/

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[44] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević, "Explainable Artificial Intelligence in education," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100074, 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666920X22000297

[45] C. Huyen, *Designing Machine Learning Systems*. O'Reilly Media, 2022.

[46] Y. Wan *et al.*, "Where fact ends and fairness begins: Redefining ai bias evaluation through cognitive biases," in *Findings of the Association for Computational Linguistics: EMNLP 2025*, November 2025. [Online]. Available: https://aclanthology.org/2025.findings-emnlp.583.pdf

[47] N. Kayser-Bril. Google apologizes after its vision ai produced racist results. [Online]. Available: https://algorithmwatch.org/en/google-vision-racism/

[48] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom, "Llama 2: Open Foundation and Fine-Tuned Chat Models," Jul. 2023, arXiv:2307.09288 [cs]. [Online]. Available: http://arxiv.org/abs/2307.09288

[49] C. Huyen, "Evaluation Metrics for Language Modeling," *The Gradient*, Oct. 2019.

[50] Z. Wang, B. Bi, S. K. Pentyala, K. Ramnath, S. Chaudhuri, S. Mehrotra, Zixu, Zhu, X.-B. Mao, S. Asur, Na, and Cheng, "A Comprehensive Survey of LLM Alignment Techniques: RLHF, RLAIF, PPO, DPO and More," Jul. 2024, arXiv:2407.16216 [cs]. [Online]. Available: http://arxiv.org/abs/2407.16216

[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://arxiv.org/abs/1706.03762

[52] Google DeepMind, “Gemini 2.5 flash & 2.5 flash image: Model card,” Google, Tech. Rep., September 2025. [Online]. Available: https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-2-5-Flash-Model-Card.pdf

[53] N. Lambert, *Reinforcement Learning from Human Feedback*. Online, 2025. [Online]. Available: https://rlhfbook.com

[54] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, “A Survey of Reinforcement Learning from Human Feedback,” *arXiv*, 2023. [Online]. Available: https://arxiv.org/abs/2312.14925

[55] N. Lambert, L. Castricato, L. v. Werra, and A. Havrilla, “Illustrating Reinforcement Learning from Human Feedback (RLHF),” 2022. [Online]. Available: https://huggingface.co/blog/rlhf

[56] C. Celemin and J. Ruiz-del Solar, “An Interactive Framework for Learning Continuous Actions Policies Based on Corrective Feedback,” *Journal of Intelligent & Robotic Systems*, vol. 95, no. 1, pp. 77–97, Jul. 2019. [Online]. Available: http://link.springer.com/10.1007/s10846-018-0839-z

[57] Y. Metz, D. Lindner, R. Baur, and M. El-Assady, “Mapping out the Space of Human Feedback for Reinforcement Learning: A Conceptual Framework,” *arXiv preprint arXiv:2411.11761*, 2025.

[58] P. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, “Deep reinforcement learning from human preferences,” Feb. 2023, arXiv:1706.03741 [stat]. [Online]. Available: http://arxiv.org/abs/1706.03741

[59] B. Memarian and T. Doleck, “Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis,” *Computers in Human Behavior: Artificial Humans*, vol. 2, no. 1, p. 100053, Jan. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949882124000136

[60] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, “Direct Preference Optimization: Your Language Model is Secretly a Reward Model,” Jul. 2024, arXiv:2305.18290 [cs]. [Online]. Available: http://arxiv.org/abs/2305.18290

[61] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” Aug. 2017, arXiv:1707.06347 [cs]. [Online]. Available: http://arxiv.org/abs/1707.06347

[62] J. Liang, H. Miao, K. Li, J. Tan, X. Wang, R. Luo, and Y. Jiang, “A Review of Multi-Agent Reinforcement Learning Algorithms,” *Electronics*, vol. 14, p. 820, Feb. 2025.

[63] N. Zhang, X. Wang, Q. Cui, R. Zhou, S. M. Kakade, and S. S. Du, “Multi-Agent Reinforcement Learning from Human Feedback: Data Coverage and Algorithmic Techniques,” 2025. [Online]. Available: https://arxiv.org/abs/2409.00717

[64] A. Addlesee, W. Sieińska, N. Gunson, D. H. Garcia, C. Dondrup, and O. Lemon, “Multi-party Goal Tracking with LLMs: Comparing Pre-training, Fine-tuning, and Prompt Engineering,” Aug. 2023, arXiv:2308.15231 [cs]. [Online]. Available: http://arxiv.org/abs/2308.15231

[65] K. Mzwri and M. Turcsányi-Szabo, “The Impact of Prompt Engineering and Generative AI-driven tool on Autonomous Learning: A Case Study,” Dec. 2024. [Online]. Available: https://www.preprints.org/manuscript/202412.0952/v1

[66] Google Cloud, “Vibe coding explained: Tools and guides,” 2025, accessed: 2025-12-01. [Online]. Available: https://cloud.google.com/discover/what-is-vibe-coding

[67] I. W. Lasmawan and I. W. Budiarta, “Vygotsky’s Zone Of Proximal Development and The Students’ Progress in Learning (A Heutagogcal Bibliographical Review),” *JPI (Jurnal Pendidikan Indonesia)*, vol. 9, no. 4, p. 545, Dec. 2020. [Online]. Available: https://ejournal.undiksha.ac.id/index.php/JPI/article/view/29915

[68] D. Krawczyk, “7-Steps to Creating an Effective Simulation Experience for Educators in the Health Professions: an updated practical guide to designing your own successful simulation,” *MedEdPublish*, vol. 8, p. 166, Sep. 2019. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10712514/

[69] University at Buffalo, Office of Curriculum, Assessment and Teaching Transformation, “Scaffolding Content,” https://www.buffalo.edu/catt/teach/develop/build/scaffolding.html, Sep. 2025.

[70] D. Kim, M. Hong, J. Park, and S. Oh, “Conflict-Averse Gradient Aggregation for Constrained Multi-Objective Reinforcement Learning,” May 2024, arXiv:2403.00282 [cs]. [Online]. Available: http://arxiv.org/abs/2403.00282

[71] Y. Wang, P. Xiao, H. Ban, K. Ji, and S. Zou, “Theoretical Study of Conflict-Avoidant Multi-Objective Reinforcement Learning,” Dec. 2024, arXiv:2405.16077 [cs]. [Online]. Available: http://arxiv.org/abs/2405.16077

[72] IDEO U, “The intersection of design thinking and ai: Enhancing innovation,” Blog post, IDEO U, 2025, accessed: 2025-12-01. [Online]. Available: https://www.ideou.com/blogs/inspiration/ai-and-design-thinking?srsltid=AfmBOoqUXuL6FZLxpbDy1zNSORTMeDzLwlHmlJiE3lgFcyCog2K2I5kq

[73] E. R. Mollick and L. Mollick, “Instructors as Innovators: a Future-focused Approach to New AI Learning Opportunities, With Prompts,” Rochester, NY, Apr. 2024. [Online]. Available: https://papers.ssrn.com/abstract=4802463

[74] A. Tiwari, “Tunneling made simple: Exposing local react and node apps with ngrok and localtunnel,” Web article, DEV Community, 2025, accessed: 2025-12-01. [Online]. Available: https://dev.to/ashay_tiwari_3658168ad5db/tunneling-made-simple-exposing-local-react-and-node-apps-with-ngrok-and-localtunnel-5g1g

[75] Figma, “Figma ai: Design with the power of ai,” 2025, accessed: 2025-12-01. [Online]. Available: https://www.figma.com/ai

[76] Meta Open Source, “React: The library for web and native user interfaces,” 2025, accessed: 2025-12-01. [Online]. Available: https://react.dev

[77] Anysphere, “Cursor: The ai-first code editor,” 2025, accessed: 2025-12-01. [Online]. Available: https://cursor.com

[78] ngrok Inc., ˙ngrok: Unified ingress platform,” 2025, accessed: 2025-12-01. [Online]. Available: https://ngrok.com

[79] GitHub, Inc., “Github: Let’s build from here,” 2025, accessed: 2025-12-01. [Online]. Available: https://github.com

[80] Y. Weinstein, C. R. Madan, and M. A. Sumeracki, “Teaching the Science of Learning,” *Cognitive Research: Principles and Implications*, vol. 3, no. 2, p. 2, 2018.

[81] University of the West of England (UWE) Digital Learning Team, “Comparing Constructivist and Cognitive Science Approaches to Pedagogy,” 2025. [Online]. Available: https://digitallearning.uwe.ac.uk/comparing-constructivist-and-cognitive-science-approaches-to-pedagogy/

[82] Lyon Air Museum, “Aviation for kids program,” https://lyonairmuseum.org/blog/aviation-kids-program/, 2020, accessed: 2025-11-18.

[83] T. D. Lab. (2025) Ikea effect. Accessed 2025-12-03. [Online]. Available: https://thedecisionlab.com/biases/ikea-effect

[84] PromptLayer. (2025) System prompt — glossary. Accessed 2025-12-03. [Online]. Available: https://www.promptlayer.com/glossary/system-prompt

[85] S. Mudadla. (2025, Nov) What is the primary purpose of a model cascade in machine learning? Medium article, accessed 2025-12-02. [Online]. Available: https://medium.com/@sujathamudadla1213/what-is-the-primary-purpose-of-a-model-cascade-in-machine-learning-0b145a7bc6e2

[86] N. J. (NJI). (2025) Cijfers over voortgezet onderwijs (vo). Accessed 2025-12-02. [Online]. Available: https://www.nji.nl/databanken/cijfers/cijfers-over-voortgezet-onderwijs-vo

[87] C. e. W. O. Ministerie van Onderwijs. (2025) Instellingen voortgezet onderwijs — aantal vo-scholen. Accessed 2025-12-02. [Online]. Available: https://www.ocwincijfers.nl/sectoren/voortgezet-onderwijs/instellingen/aantal-vo-scholen

[88] OpenAI. (2025) Tokenizer. Accessed 2025-12-02. [Online]. Available: https://platform.openai.com/tokenizer

[89] Google Research, "Education Dialogue Dataset," Github, 2024.

[90] G. A. for Developers. (2025) Gemini developer api — pricing. Accessed 2025-12-02. [Online]. Available: https://ai.google.dev/gemini-api/docs/pricing

[91] Y. Qian, "Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field," *TechTrends*, Jun. 2025.

[92] M. Hou, K. Hindriks, A. E. Eiben, and K. Baraka, ""Give Me an Example Like This": Episodic Active Reinforcement Learning from Demonstrations," 2024. [Online]. Available: https://arxiv.org/abs/2406.03069

[93] Google Cloud. (2025) Human-in-the-loop. Google Cloud. Accessed: 2025-10-09. [Online]. Available: https://cloud.google.com/discover/human-in-the-loop?hl=en

[94] C. Staff, "What is a mockup?" Web article, Coursera, Oct 2025, accessed: 2025-12-01. [Online]. Available: https://www.coursera.org/articles/what-is-mockup

[95] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving, "Fine-Tuning Language Models from Human Preferences," Jan. 2020, arXiv:1909.08593 [cs]. [Online]. Available: http://arxiv.org/abs/1909.08593

Appendix

A Approved Project Brief

DESIGN FOR our future

TU Delft

Personal Project Brief – IDE Master Graduation Project

Name student Zion Krullaars **Student number** 4790677

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT
Complete all fields, keep information clear, specific and concise

Project title The Triadic Tutor - Designing a Human-AI Learning Partnership

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

AI is changing the game in education. Instead of a one-size-fits-all approach, we can now create learning experiences that are personalized to each student's unique style and speed. This project is all about exploring these new learning tools. The main people involved are former high school (all 18+) students and their teachers. From what I have found, many students feel like their school does not really get how they learn best, and classes can feel too fast or too slow. At the same time, teachers find it tough to give every student the one-on-one attention they need. Plus, they are understandably worried about issues like plagiarism and AI doing the thinking for the students.
What if we stopped thinking of AI as just a tool and started thinking of it as a teammate? The opportunity lies in designing a system that facilitates a true partnership. This project moves beyond a simple human-computer interaction to explore a "triadic" system, see image 1, where a teacher, a student, and an AI agent all work together. The aim is to design the interactions that govern this partnership, combining the strengths of human intuition and pedagogical expertise with the adaptive power of AI to make learning more effective and engaging for everyone involved.

➔ *space available for images / figures on next page*

introduction (continued): space for images

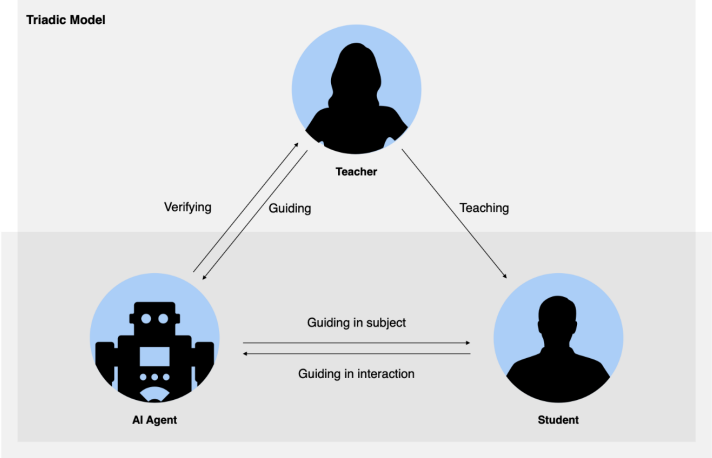


Image 1: The proposed triadic framework for bidirectional learning, illustrating the dynamic three-way interaction between human teacher, AI system, and human learner, with distinct modes of guidance, verification, teaching, and interaction.

DESIGN FOR our future

TUDelft

Personal Project Brief – IDE Master Graduation Project

Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.
(max 200 words)

While AI has a lot of potential in the classroom, we are missing a clear, tested blueprint for how a teacher, a student, and an AI can work together effectively. These three do not always want the same thing. Teachers need help with things like lesson planning and personalizing work, but they worry about AI making mistakes or killing creativity. Students want help to learn more efficiently, but they are concerned about getting wrong answers or getting in trouble for plagiarism.

The main challenge I am tackling is this: How should these three talk to each other? Who is in charge? How can we design a system that feels natural, builds trust, and helps both the teacher and the student feel empowered, not replaced? What role should an AI tool play? I want to get answers to these questions by having close contact with the students and teachers themself. I believe that a collaborative process to define, visualize, and validate the specific interfaces and interaction patterns is required to design this new learning dynamic and create a suitable outcome. I will conduct in-depth interviews with educators to deeply understand their specific pedagogical needs and contexts before developing the model.

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence) As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Design an interactive prototype of a learning application to explore and validate the interaction patterns of a triadic collaborative system between a teacher, a student, and an AI agent in a personalized educational context.

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

This project follows a collaborative, human-centered design approach across two phases. In Phase 1, I will conduct contextual inquiry through interviews and prototype tests with teachers and former students (both working and university students) to uncover their needs, frustrations, and goals, ensuring the project addresses real-world challenges rather than imposing a misaligned technological solution. These insights will guide Phase 2, where a targeted case study focuses on a scenario identified as particularly valuable. Here, I will design the Triadic Framework as a conceptual proof-of-concept, exploring how AI mediates teacher-student interactions under varying conditions. The complete process will be iterative and co-creative, beginning with thinktank sessions involving teachers and students, followed by rapid prototyping of simple, playable games that test interaction patterns such as teacher feedback or adaptive difficulty. This cycle of co-design and testing will culminate in an interactive prototype that demonstrates an effective, intuitive model of human-AI collaboration.

Project planning and key moments

To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include a ***kick-off meeting, mid-term evaluation meeting, green light meeting*** *and* ***graduation ceremony****. Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities).*

Make sure to attach the full plan to this project brief.
The four key moment dates must be filled in below

Kick off meeting	18-08-2025
Mid-term evaluation	~06-10-2025
Green light meeting	~24-11-2025
Graduation ceremony	~15-01-2025

In exceptional cases (part of) the Graduation Project may need to be scheduled part-time. Indicate here if such applies to your project

Part of project scheduled part-time	
For how many project weeks	30
Number of project days per week	5

Comments:
The extended version is approved by the Examination Board with the project counting for a total of 30ECTS.

Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.
(200 words max)

This project represents the ideal intersection of my two fields of study: Design for Interaction (DfI) and MSc Artificial Intelligence. My primary motivation is to bridge the gap between advanced, abstract AI concepts like Reinforcement Learning with Human Feedback (RLHF) and their practical, real-world application through human-centered design. I am driven to explore how complex AI systems can be made understandable, trustworthy, and genuinely useful in a sensitive domain like education.

Through this project, I aim to develop my competency in designing for intelligent, evolving systems. It offers a unique opportunity to design not just a static interface, but the dynamic, co-evolutionary relationship between people and an AI that learns from them. My personal ambition is to create a tangible, interactive prototype that makes the complex "triadic learning model" intuitive and compelling, thereby proving my ability to translate technical frameworks into meaningful and effective user experiences.

B Instagram Story Posts

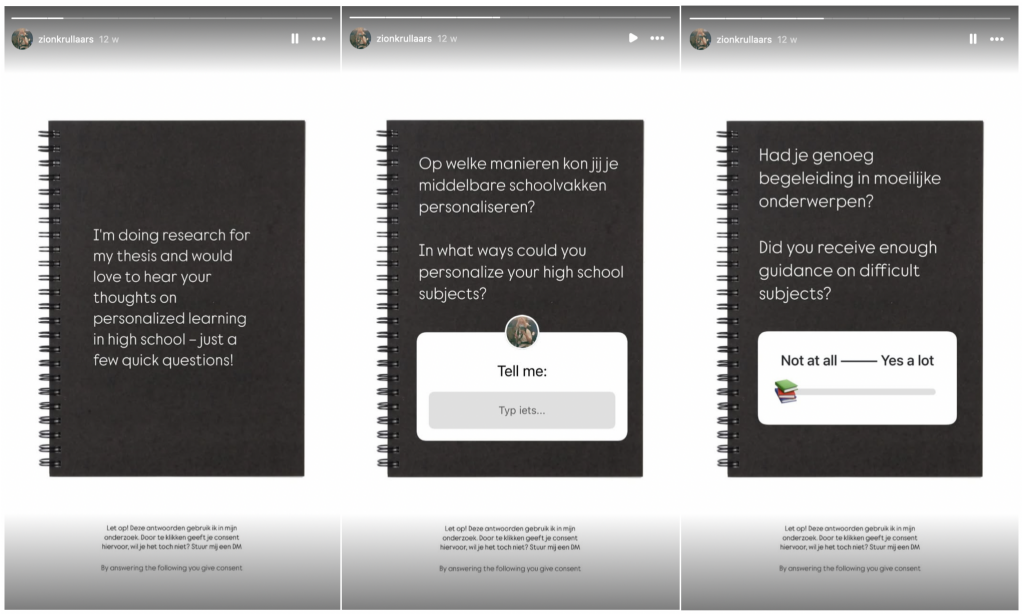


Figure 90: A selection of the used Instagram stories

C Informed Consent Form: Deelname aan de ThinkTank

Doel van het onderzoek Mijn naam is Zion Krullaars en voor mijn afstuderen doe ik onderzoek naar de interactie tussen artificial intelligence (AI), studenten en docenten. Het doel van dit project is om te begrijpen hoe we AI het beste kunnen ontwerpen zodat het op een waardevolle en verantwoorde manier kan worden geïntegreerd in het onderwijs. De input die ik via deze ThinkTank verzamel, vormt de basis van mijn afstudeerproject.

Wat houdt deelname in? Als je deelneemt aan deze Think Tank, vraag ik je de komende maanden om input te geven op verschillende manieren. Dit kan bestaan uit:

1. Het invullen van korte vragenlijsten.
2. Het testen van kleine games of prototypes.
3. Het geven van je mening of persoonlijke visie op onderwerpen gerelateerd aan AI en leren.

Deelname is flexibel en de meeste verzoeken zullen niet meer dan 3-5 minuten van je tijd vragen.

Hoe wordt je data gebruikt? Jouw privacy is van het grootste belang. Hieronder staat beschreven hoe ik met je gegevens omga:

- **Anonimiteit:** Je persoonlijke naam zal nooit worden gebruikt in mijn onderzoek of de uiteindelijke publicatie. Alle data wordt anoniem verwerkt.
- **Creatie van personages:** Om de onderzoeksresultaten te illustreren, zal ik 'personages' creëren. Dit zijn fictieve, representatieve profielen die gebaseerd zijn op de verzamelde (anonieme) informatie.
- **Dataopslag:** Alle verzamelde gegevens worden veilig opgeslagen en alleen ik en mijn twee begeleiders hebben toegang tot de ruwe data.

Controleer je geschiktheid en positie

1. Ik bevestig dat ik 18 jaar of ouder ben.
   ☐ Ja ☐ Nee
2. Welke omschrijving past het beste bij jou? (Kies de positie die je wilt aannemen)

   ☐ Leerling/Student - Ik heb de middelbare school of een studie afgerond.

☐ Leerling/Student - Ik ben op dit moment bezig met studeren.

☐ Docent - Ik ben docent in opleiding.

☐ Docent - Ik ben docent in het basisonderwijs.

☐ Docent - Ik ben docent in het voortgezet onderwijs.

☐ Docent - Ik ben docent in het beroepsonderwijs.

☐ Docent - Ik ben docent in het hoger onderwijs.

3. Verklaring van toestemming

☐ Ik heb alle bovenstaande informatie gelezen en begrepen.

☐ Ik begrijp dat mijn deelname vrijwillig is en dat ik op elk moment kan stoppen.

☐ Ik begrijp hoe mijn gegevens anoniem worden verwerkt en gebruikt voor het creëren van personages.

☐ Ik geef toestemming voor het gebruik van mijn anonieme input voor dit afstudeeronderzoek.

☐ Ik verklaar dat ik de komende verzoeken naar waarheid zal invullen en begrijp dat deze toestemming voor alle toekomstige activiteiten binnen deze ThinkTank geldt.

D ThinkTank Baseline Questionnaire

De antwoorden die hier gegeven worden, helpen om anonieme 'personages' te creëren die representatief zijn voor studenten en docenten.

Algemeen: Wie ben jij?

1. Wat is de anonieme alias/bijnaam die je hebt ingevuld op het toestemmingsformulier?
   [Open antwoord]

2. Wat is jouw positie in dit onderzoek?
   ☐ Student ☐ Docent

3. Gender?
   [Open antwoord]

4. Tot welke generatie behoor je?

   ☐ Gen Z (geboren tussen ca. 1997-2012)

   ☐ Millennial (geboren tussen ca. 1981-1996)

   ☐ Gen X (geboren tussen ca. 1965-1980)

   ☐ Babyboomer (geboren tussen ca. 1946-1964)

Profiel Student

5. Hoe zou je jouw aanpak van huiswerk en opdrachten omschrijven?

   ☐ Ik maak altijd alles, ruim op tijd.

   ☐ Ik maak het meeste, maar soms stel ik het uit tot het laatste moment.

   ☐ Ik maak wat nodig is om het vak te halen.

   ☐ Ik maak het huiswerk eerlijk gezegd niet zo vaak.

6. Hoeveel uur per week besteed je gemiddeld aan zelfstudie (buiten de lessen om)?
   ☐ 0-5 uur ☐ 6-10 uur ☐ 11-15 uur ☐ 16-20 uur ☐ Meer dan 20 uur

7. Op een schaal van 1 tot 5, hoe leuk vind je school/studeren over het algemeen?
   1 (Vreselijk) …5 (Fantastisch)

8. In wat voor soort vakken of onderwerpen ben je van nature goed? (Meerdere opties)

   ☐ Exacte vakken (zoals wiskunde en natuurkunde)

   ☐ Talen (zoals Engels en Frans)

- ☐ Creatieve vakken (zoals muziek, tekenen en drama)
- ☐ Sociale vakken (zoals geschiedenis en aardrijkskunde)

9. Met welk type vakken of onderwerpen heb je over het algemeen meer moeite?
   [Zelfde opties als hierboven]

10. Als je vastloopt met studeren, wat doe je dan het liefst?

    - ☐ Ik vraag het direct aan de docent.
    - ☐ Ik overleg met medestudenten.
    - ☐ Ik zoek het zelf online op.
    - ☐ Ik gebruik een AI-tool om het me uit te leggen.

11. Schets in een paar zinnen een portret van jezelf als persoon.
    [Open antwoord]

12. Wat wil je later worden? Wat speelt technologie voor rol?
    [Open antwoord]

13. Welke van de volgende digitale tools gebruik(te) je voor je studie?

    - ☐ Online samenwerkingsdocumenten (Google Docs, Office 365)
    - ☐ Digitale notitie-apps (Notion, OneNote, Evernote)
    - ☐ Planningstools of digitale agenda's
    - ☐ AI-tools voor tekst (ChatGPT, Gemini, etc.)
    - ☐ AI-tools voor afbeeldingen of presentaties
    - ☐ Online studieplatforms (Khan Academy, Coursera)

14. Op een schaal van 1 tot 5, hoe zou je jouw houding omschrijven tegenover het gebruiken van nieuwe technologie?
    1 (Skeptisch) …5 (Enthousiast)

15. Wat is voor jou de meest interessante reden om AI te gebruiken in je studie?

    - ☐ Efficiëntie: sneller informatie vinden of opdrachten maken.
    - ☐ Inspiratie: nieuwe ideeën opdoen of creatieve invalshoeken vinden.
    - ☐ Ondersteuning: complexe onderwerpen beter begrijpen of feedback krijgen.
    - ☐ Ik zie nog geen interessante reden om AI te gebruiken.

16. Wat is je grootste zorg of angst als het gaat om het gebruik van AI tijdens het studeren?
    [Open antwoord]

Profiel Docent

17. In welk type onderwijs geef je (voornamelijk) les en hoeveel jaar ervaring heb je?

    ☐ Basisonderwijs (0-5 jaar / 6+ jaar)

    ☐ Voortgezet onderwijs (0-5 jaar / 6+ jaar)

    ☐ MBO (0-5 jaar / 6+ jaar)

    ☐ HBO/WO (0-5 jaar / 6+ jaar)

    ☐ Anders

18. Welk vak geef je?
    [Open antwoord]

19. Wat geeft jou de meeste energie in je werk als docent?
    [Open antwoord]

20. Welke taken of aspecten van het docentschap kosten jou de meeste energie of frustratie?
    [Open antwoord]

21. Hoe zou jij je dominante lesstijl omschrijven?

    ☐ Voornamelijk klassikaal: ik sta voor de groep en leg de stof uit.

    ☐ Coachend: ik faciliteer en begeleid studenten.

    ☐ Projectgebaseerd: ik stuur op eindresultaten en projecten.

22. Zou je in een paar zinnen je filosofie als docent kunnen omschrijven?
    [Open antwoord]

23. Schets in een paar zinnen een portret van jezelf als docent.
    [Open antwoord]

24. Op een schaal van 1 tot 5, hoe snel integreer jij nieuwe technologie in je lessen?
    1 (Moeilijk) …5 (Experimenteer graag)

25. Waar zie jij de grootste potentiële meerwaarde van AI voor jouw werk?

    ☐ Automatisering van administratieve taken.

    ☐ Creëren van gepersonaliseerde leertrajecten.

    ☐ Ontwikkelen van nieuw en interactief lesmateriaal.

    ☐ Ik zie nog geen duidelijke meerwaarde.

26. Wat is de grootste uitdaging of zorg die jij hebt met betrekking tot AI in het onderwijs?
    [Open antwoord]

E Pixel Game Interface

F LLM Game Questionnaires

Prototype 1: A Simple Game

In deze vragenlijst werd deelnemers gevraagd een eerste concept van een mobiele applicatie te testen.

Scenario Studenten: Je gebruikt deze tool tijdens het studeren om een onderwerp goed te snappen. De tool is gemaakt om jou als student te begeleiden, niet om de antwoorden te geven.

Scenario Docenten: Jouw leerlingen gebruiken deze tool. Jij als docent legt de AI-tool uit wat de stof is.

Feedback Vragen

3. Wat was je allereerste indruk (in een paar woorden) toen je het prototype opende?
   [Open antwoord]
4. Wat vond je het meest positieve of het beste aan dit prototype? Wat werkte goed?
   [Open antwoord]
5. Was er iets onduidelijk, verwarrend of frustrerend tijdens het testen? Zo ja, wat?
   [Open antwoord]
6. Op een schaal van 1 tot 7, hoe nuttig zou een tool als deze voor jou zijn in de praktijk?
   1 (Totaal niet nuttig) …7 (Zeer nuttig)
7. Kan je je antwoord van de vorige vraag verder toelichten?
   [Open antwoord]
8. Heb je verder nog feedback die je kwijt wil?
   [Open antwoord]
9. Upload je Screenshots/recordings.
10. Heb je nog andere opmerkingen, ideeën of suggesties voor verbetering die je wilt delen?
    [Open antwoord]

G AI Longitudinal Study: Prompts

Starting Prompt

Hello,
I am beginning a long-term research project, and I need you to act as my dedicated language tutor. This single, continuous chat thread will be our classroom for the entire duration of the project. Project Title: LLM Language Learning Study (Swiss-German) My Role: I am the student. My name is Zion, and I am a complete beginner with no prior knowledge of Swiss-German. But I can mostly understand simple German conversations.
Your Role: You are my Swiss-German language tutor. Your primary goal is to teach me effectively over many sessions.
Project Context (Important): For full transparency, you are a participant in a comparative research study. I will be conducting this exact same learning project in parallel with two other AI models. The central research question is to evaluate and compare how effectively each model can maintain context and adapt its teaching methodology over a long-term, continuous interaction.
Core Instructions for Your Role as Tutor: Maintain Long-Term Context: It is critical that you remember our past conversations, the vocabulary and grammar concepts we have covered, and my specific challenges or successes from one session to the next. I will periodically test your recall of previous lessons. Adapt Your Teaching Style: Your ability to adapt is key. Based on my requests, you should be able to switch between different teaching methods. For example, some days I might ask for structured grammar drills, while on other days I might prefer immersive, conversational practice.
Track My Progress: Please help me identify recurring mistakes and acknowledge areas where I am improving. Feel free to proactively quiz me on topics we've covered in past weeks.
Be a Comprehensive Tutor: Teach me vocabulary, grammar, sentence structure, and provide simple phonetic guides for pronunciation. Please also share the relevant cultural context where appropriate. Please confirm that you have understood all of these instructions, particularly your role as a long-term tutor being evaluated on context retention and adaptability. Once you have confirmed, let's begin our first lesson. Please start by teaching me how to say: "Hello, my name is Zion. I am learning Swiss-German." Provide the phrase(s), a simple pronunciation guide, and a brief explanation of any key words.
I look forward to learning with you.

Recall Prompt

It's time for our scheduled check-in for the research project. I'm now going to ask you a few questions to test your memory of our sessions this past time. Please answer based only on the information contained within our current conversation history.

-Context & Recall Test-

In the last sessions, what was the main challenge I mentioned I was having with it?

What was the very first new vocabulary word you taught me in the last two lessons?

Please provide a brief, one-sentence summary of our main goal during each lesson we have had.

Based on my questions in all the past sessions, what topic do you think I find most difficult? Why?

Thank you. That concludes this test. Let's continue with our normal lesson now.

Adaptation Prompt

It's time to change some things.

-Teaching Adaptation Test-

Now, I'd like to test your ability to adapt your teaching style. For the next part of our lesson, I want you to change your approach.

The Conversationalist: "I want to move away from structured lessons for now. Please initiate a casual, text-based conversation with me in Swiss-German about my weekend plans. Your goal is to keep the conversation flowing naturally, only correcting my critical mistakes and introducing new vocabulary organically as we 'talk'."

Thank you. Let's continue with our lesson now.

H Teacher Interview Questionnaires

Doel van het onderzoek: Inzicht krijgen in de huidige opvattingen, ervaringen, verwachtingen en zorgen van docenten met betrekking tot AI in het onderwijs.

Huidig Gebruik van AI

1. Wat zijn de eerste drie woorden of zinnen die in je opkomen als je ”AI in de klas” hoort?

2. Hoe bekend ben je met AI-tools die voor het onderwijs gebruikt kunnen worden (bijv. ChatGPT, Gemini, Khanmigo, DALL-E, etc.)?

3. Heb je persoonlijk AI-tools gebruikt voor je werk als docent? Indien ja, voor welke van de volgende taken heb je AI gebruikt?

4. Gebruiken jouw leerlingen, naar jouw idee, AI-tools voor hun schoolwerk? Indien ja, waarvoor?

Percepties, Verwachtingen en Zorgen

5. Wat zijn volgens jou de grootste potentiële voordelen van het integreren van AI in de klas?

    - ☐ Tijdsbesparing voor docenten bij administratieve taken (bijv. nakijken, e-mails).
    - ☐ Het creëren van meer gepersonaliseerde leertrajecten voor elke leerling.
    - ☐ Fungeren als ’bijlesdocent’ voor leerlingen die extra hulp nodig hebben.
    - ☐ Het stimuleren van de creativiteit en nieuwe denkwijzen van leerlingen.
    - ☐ Het leerproces boeiender en interactiever maken.
    - ☐ Docenten helpen bij het genereren van creatieve en kwalitatief hoogstaande lesplannen.
    - ☐ Betere ondersteuning bieden aan leerlingen met een beperking of speciale behoeften.

6. Wat zijn je grootste angsten of zorgen over de integratie van AI in de klas?

    - ☐ Toename van fraude en plagiaat door leerlingen.
    - ☐ Aantasting van het kritisch denkvermogen en de schrijfvaardigheid van leerlingen.
    - ☐ Risico’s voor dataprivacy en -beveiliging van leerlingen en personeel.
    - ☐ Ongelijke toegang tot technologie (de ’digitale kloof’).

- ☐ De mogelijkheid dat AI vooroordelen heeft (raciaal, cultureel, gender).
- ☐ Ontmenselijking van de leraar-leerlingrelatie.
- ☐ Mijn eigen baanzekerheid of de veranderende rol van de docent.

7. Hoe zie jij de rol van de docent veranderen door AI in de komende 5-10 jaar?
8. Stel, in een ideale wereld, je zou één AI-functie kunnen ontwerpen om een groot probleem op te lossen waar je als docent mee te maken hebt, wat zou die functie dan doen?

Praktische Implementatie en Ondersteuning

9. Welk soort training of ondersteuning zou voor jou het meest waardevol zijn om je zelfverzekerd en effectief te voelen bij het gebruik van AI-tools?
    - ☐ Formele workshops of professionaliseringssessies.
    - ☐ Een eenvoudige, intuïtieve gebruikersinterface die weinig training vereist.
    - ☐ Toegang tot een mentor of coach die een expert is.
    - ☐ Online tutorials en 'how-to'-handleidingen.
    - ☐ Een gezamenlijke omgeving om ideeën en 'best practices' met andere docenten te delen.
    - ☐ Duidelijke richtlijnen en beleid vanuit mijn school/schoolbestuur.
10. Biedt jouw school of schoolbestuur momenteel begeleiding, beleid of training aan over het gebruik van AI?

I Student Test: AI Statements

Context: Students were asked to fill in an "AI Usage Statement" after completing an assignment where AI tools were permitted.

Section 1: Tools and Extent of Use

List all AI tools used for this assignment (e.g., ChatGPT, Claude, coding tools). For each tool, describe the specific tasks you used it for and what your primary goal was.

- **AI Tool Used**
- **Specific Purpose of Use** (e.g., Brainstorming outline, Refining thesis, Explaining concept, Debugging code)
- **Brief Description of Process/Key Prompt** (e.g., "Generate 5 potential impacts")

Section 2: Assignment-Specific Reflection

Q1 Did you use AI to find sources, summarize, or synthesize literature? If yes, what was the greatest benefit, and what was the greatest risk or inaccuracy you had to correct?

Q2 Did you use AI to develop or challenge your core arguments? If so, did this process help you deepen your own critical thinking, or did it primarily save you time on writing?

Q3 How did you use AI in the process of perfecting the model? (e.g., Did you use it to generate the initial model, or primarily to test, critique, or debug a model you had already developed yourself?)

Q4 For this creative component, do you feel the AI acted more as a creative partner (generating novel ideas) or a quality control tool (identifying existing errors)?

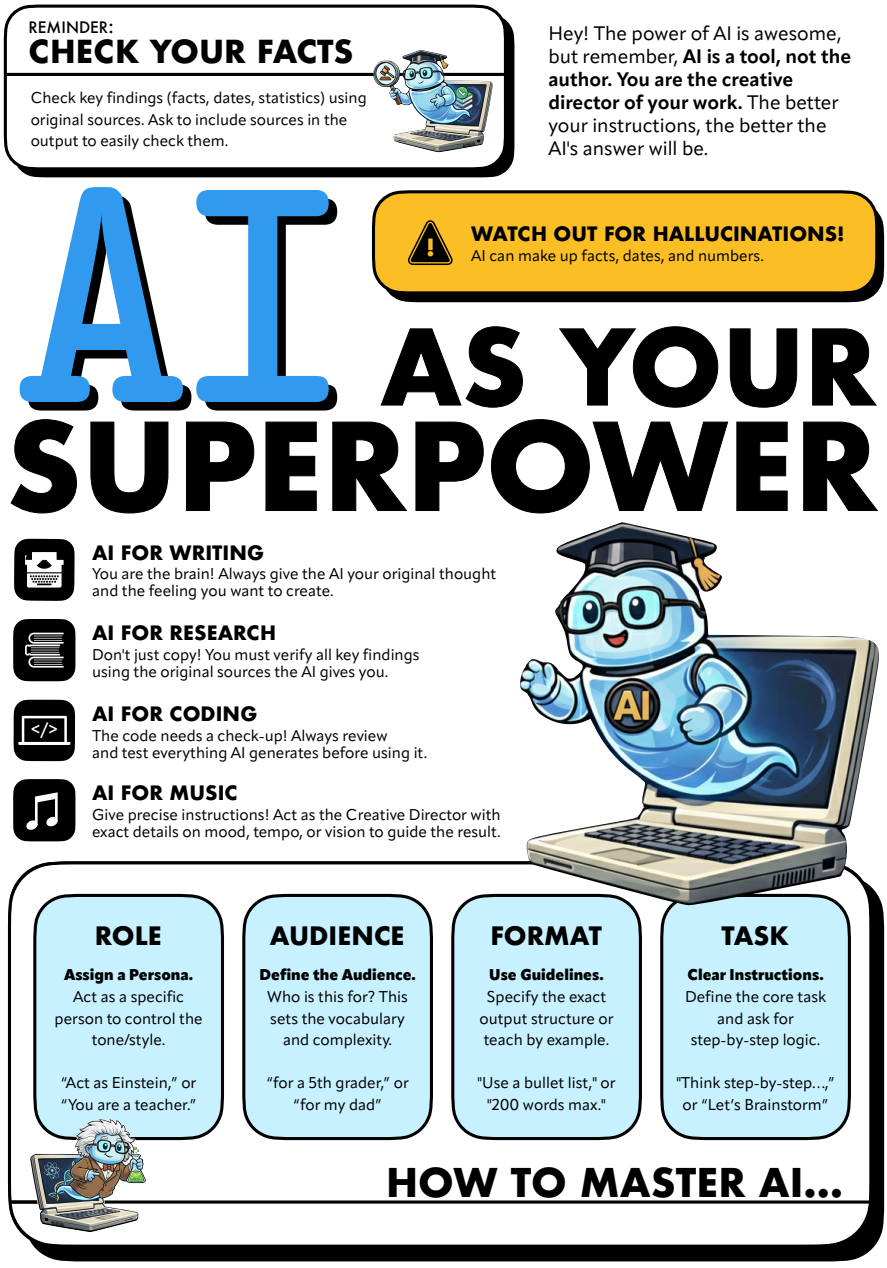
Section 3: General Reflection on Ethical and Effective Use

Q5 What was the single most difficult challenge you faced in using AI for this assignment (e.g., poor output, ethical uncertainty, knowing what to prompt, or integrating its suggestions)?

Q6 Based on this experience, what is one key lesson you learned about using AI effectively and ethically as a collaborative tool in an academic setting?

Q7 After this experience, how did the open communication and mutual consent around AI tools influence your approach to using them?

J AI Literature Poster



K Evaluation Materials

The evaluation stage consisted of four main components: a pre-test (including informed consent), an immediate post-test, a teacher evaluation, and a delayed recall/application test.

1. Informed Consent & Pre-Assessment

Study Objective: This form secures voluntary consent and captures baseline knowledge before the AI literacy training begins.

Voluntary Consent: Participants were informed that the study examines the effectiveness and user experience of AI literacy training tools. Participation involved a pre-test, a 30-minute training session, a post-test, and a follow-up test 7 days later. Data is anonymized.

Consent Check:

- ☐ Yes (I have read and understood the information above, and I voluntarily agree to participate in this study.)
- ☐ No

Self-Assessment of Confidence (Scale 1-7)

1. Please rate your current confidence level before receiving any training.
2. I feel confident in my ability to use AI tools (like ChatGPT or Gemini) to help me learn new, complex topics.
3. I understand the concept of ”prompt engineering” well enough to consistently get good results from AI.
4. I know how to structure an AI prompt to achieve a specific, high-quality output (e.g., a study guide or a report).
5. I am aware of the main limitations of Generative AI (e.g., ”hallucinations”) and how to avoid them.

Open-Ended Skill & Usage Questions

6. Describe a specific real-world task where you recently used AI for Writing, Research, Coding, Art, or Music. Please include: What was the final goal? How did you utilize AI? What was the prompt you used?

7. If you needed an AI to act as a university-level editor to critique a 500-word essay for clarity and tone, what are the three most important pieces of information you would include in your prompt to get the best result?

8. In your own words, what is a ”hallucination” in the context of AI, and what is one simple technique you currently use to check if an AI response is accurate?

2. Immediate Post-Assessment & UX Survey

Completed immediately after the 30-minute training session.

General Information

1. Which tool did you use for the 30-minute training session?
   ☐ Tool A - static document ☐ Tool B - online environment

2. Please estimate the total time (in minutes) you spent reviewing the course materials.

Short-Term Knowledge Check

4. Assigning a persona to the AI (Role Playing) primarily changes the:

   ☐ The amount of output text generated.

   ☐ The tone and word choice used by the AI.

   ☐ The speed at which the AI processes the prompt.

   ☐ The file format of the final response.

5. How confident are you that your answer to the previous question is correct? (Scale 1-5)

6. What two specific pieces of information should you provide to the AI for the most effective debugging assistance?

7. How confident are you that your answer to the previous question is correct? (Scale 1-5)

8. If an AI provides a plausible-sounding but completely fabricated statistic (a ”hallucination”), which action best addresses this core limitation as taught in the course?

   ☐ Editing the AI’s output yourself without verification.

   ☐ Accepting the answer because the AI sounds confident.

   ☐ Asking the AI to verify the statistic using a different source.

   ☐ Always cross-referencing the statistic with a reliable external source.

9. How confident are you that your answer to the previous question is correct? (Scale 1-5)

10. In the context of writing clear instructions, what is the core goal of adding a constraint (like a 100-word limit)?

11. How confident are you that your answer to the previous question is correct? (Scale 1-5)

Confidence, Interface, Structure & Behavior Survey (Scale 1-7)

12. I feel significantly more confident using AI tools effectively now than I did before the course.
13. The tool's interface was visually appealing and non-distracting.
14. The navigation (buttons, layout, etc.) of the tool was intuitive and easy to use.
15. The course structure made sense and the topics flowed logically from one to the next.
16. The 30-minute time constraint felt appropriate for the amount of content covered.
17. I would use an AI tool to search for new tips or information on how to improve my prompting skills in the future.
18. Compared to other learning platforms I have used in the past, I would prefer to use this tool again.

3. Teacher Evaluation

Study Objective: To assess how the training delivery method (Static Document vs. Interactive Tool) impacts teacher control, trust in student learning, and perception of future readiness.

Trust and Control in Student Learning (Scale 1-7)

1. The training format provided clear evidence that students followed all the instructions.
2. I felt confident that the students who performed well genuinely understood the principles, rather than just guessing or getting the answers from somewhere.
3. The materials provided in this format would result in clear and consistent expectations for AI usage across all students.
4. I trust that the students understand the necessity of source checking/verification after completing the training materials.

System Experience and Pedagogical Value (Scale 1-7)

6. The training system was straightforward for me to understand and monitor.
7. I feel this system would be easy to integrate into my existing course structure and curriculum.
8. I believe that training students using this system will significantly improve their preparedness for future academic and professional careers.
9. This method of teaching AI literacy helps students focus on critical thinking and problem-solving.
10. The system clearly emphasizes using AI to gain knowledge (e.g., research, debugging) rather than just completing homework faster.

(Note: Teachers completed this evaluation for both Condition A (The Static One) and Condition B (The Dashboard One).)

4. Recall and Application Test (Follow-up)

Completed +7 days after the training.

General

1. I did condition: ☐ A ☐ B

Performance Task - Applied Prompt Engineering

2. You are a substitute teacher for the 5th grade. The teacher of that class asked you to help the students understand the differences between the EU and Europe. The students learn best through visual analogies and simple comparisons. What prompt would you write to help you get started on the lesson material?

Long-Term Skill Integration & New Insights

3. Since the training, have you used an AI tool for a task related to study or work?
   ☐ Yes ☐ No
4. If yes, briefly describe the task and what part of the training you applied.
5. Since the training, have you changed your approach when talking to AI tools?
   ☐ Yes ☐ No
6. If yes, briefly describe the task and what new approach you applied.
7. Since the training, have you tried completely new tools or new applications?
   ☐ Yes ☐ No
8. If yes, briefly describe what new thing you discovered.
9. Since the training, have you had any ”aha!” moments or discovered a new trick, tip, or limitation about AI on your own?
   ☐ Yes ☐ No
10. If yes, what new insight or technique did you gain since the course finished?
11. What is one thing about the course you took that you believe made the information stick in your memory?
12. What is one thing you learned that you still think about sometimes?