Thesis Review 1 Script

Slide1 - Thesis Review 1 *Mentor (BIEL)
Good afternoon

Today we'd like to introduce you to a more detailed view of our Thesis proposal to get things started, but before we dive into it, allow us to briefly explain how we got to this idea.

Slide 2 - Context (BIEL)
Let's begin by giving you some context to our research

Slide 3 - From Theory to Practice (BIEL)
It all clicked shortly after we started our initial brief, while developing our research analysis for AI Theory during last term.

During our research about the Anthropomorphic nature of current AI tools (which is basically the capacity of exaggerating AI capabilities and performance by attributing human-like traits to systems that do not possess them). We discovered some troubling facts.

Slide 4 - AI Interaction Changes Brain Function (BIEL)

A groundbreaking neuroscience research from MIT's Media Lab provided unprecedented evidence that AI interaction fundamentally alters human brain function, by using high-density EEG monitoring to track brain activity across 54 participants writing essays with different levels of AI assistance.

We aren't just changing how we think about AI; AI is changing how we think, period.

Slide 5 - The Problem (BIEL)
The results were striking:

Slide 6 - The Reveal: Brain Damage (BIEL)
ChatGPT users showed up to 55% reduction in neural connectivity compared to unassisted writers, measured through Dynamic Directed Transfer Function (dDTF) analysis. This metric tracks information flow across brain regions, indicating executive function, semantic processing, and attention regulation.

Users progressively shifted from asking structural questions to copying entire AI-generated essays, demonstrating what researchers term "**cognitive offloading**" that weakens rather than strengthens mental capabilities.

Most concerning, these neural changes persisted even after AI assistance was removed. When heavy AI users attempted unassisted writing, they exhibited continued neural under-engagement, suggesting lasting alterations to cognitive processing patterns. Conversely, participants who began with brain-only writing maintained stronger neural connectivity even when later using AI tools, indicating that the [sequence of AI integration] critically determines its impact on human cognition.

Slide 7 - Our Test (BIEL)

To test this we conducted an experimental conversation between an architectural designer and her AI digital twin as they collaborated on redesigning an abandoned warehouse into a community center, a scenario that was deliberately chosen to examine how creative professionals could interact with AI versions of themselves.

What emerged from this interaction revealed complex psychological dynamics. The AI consistently offered sophisticated design solutions demonstrating technical competence and cultural sensitivity that might genuinely enhance the design process.

Slide 8 - The Creative Trap: Designer Meets AI Twin (BIEL)

However after a much deeper analysis of the conversation script we realized that the designer increasingly deferred to the AI's suggestions rather than engaging in the cognitive struggle necessary for authentic creative development. When the AI proposed specific solutions the designer responded with immediate acceptance rather than critical evaluation or creative synthesis. This pattern suggests what researchers identify as "**cognitive offloading**" in creative contexts, where AI assistance may actually diminish the designer's engagement with fundamental design challenges.

Most tellingly, when the AI demonstrated "memory" of previous conversations and upgraded capabilities, the designer responded with apparent satisfaction and growing trust in the artificial relationship. This anthropomorphic response, when we start treating the AI as a colleague with improving abilities rather than a tool with updated algorithms, exemplifies how even highly trained professionals risk mistaking sophisticated pattern matching for genuine creative partnership.

Slide 9 - Preventing Cognitive BAnkruptcy (BIEL)

As a result we couldn't ignore the paradoxical nature of using AI tools when such cognitive liabilities arise, precisely because it highlights the need to course correct the development of these tools if we don't want to inflict unprecedented lasting effects on future generations of professionals. In Architecture alone a research made by Gerlich unveiled the negative correlation between frequent AI usage and critical thinking abilities, which is why we found an opportunity to address this issue as part of our Thesis proposal.

Slide 10 - Motivation (BIEL)

And most of what motivates us for this project came as a result of those discoveries

Slide 11 - What We Realized (BIEL)

Since such evidence directly contradicts narratives of AI as a tool for cognitive enhancement. [Information theory analysis] reveals that when AI systems pre-process information, humans receive “compressed” knowledge without engaging in the cognitive work necessary for deep learning and neural development.

These findings suggest that AI dependency creates measurable “cognitive debt“, a weakening of neural networks that may prove difficult to reverse, which directly impacted the way in which we viewed and approached our initial Thesis hypothesis.

Slide 12 Original vision: Then Vs Now (BIEL)
Our original abstract centered around the idea of creating a multimodal AI tool capable to mentor students and professionals through their cognitive and intellectual development by implementing an ultra-focused and practice area specific knowledge base. In that scenario, our plan was to train our tool in a narrow selection of master works of Architecture while developing personality specific personas based on the works of the greatest Architectural masters of the last century in order to test how education institutions could approach the development of AI support tools for their students while maintaining academic integrity in the process.

Slide 13 - Simply Narrowing AI Selection (BIEL)
However, after the outcome of our initial research, we realized that simply narrowing the selection and tailoring the personality traits and knowledge base of these tools will likely continue to encourage the same level of cognitive offload that currently available tools facilitate, contradicting the very basis of a purposeful cognitive development that superior education and professional practice requires, which made us rethink our approach and strategy.

Slide 14 - Reducing Cognitive Offloading (BIEL)
Then Reducing Cognitive offloading became part of our mission in this process. And our first approach would be by developing a scalable mentorship system that is capable of enforcing a constructive feedback system, where ideas are grown symbiotically together rather than just given to us. And by exploring the development of tools that think WITH us and not FOR us, we reduce the risk of shallow learning patterning.

Slide 15 - Convergence of AI (BIEL)
And since the relative recent expansion of those systems have been broadly constraint to Generative AI, Performance Evaluation and Design Automation, we wanted our thesis to revolve around a largely unexplored domain in which artificial intelligence enhances the user’s ability to understand, articulate and develop their spatial thinking (in the case of Architecture, which will be our main focus)

Slide 16 - Mentor Sits Here (BIEL)
And precisely here where we’ll position our research: Within the field of collaborative human-AI cognition synergies.

Slide 17 - Objectives (MAHNOOR)
So where would this take us?

Slide 18 - The new Mentor (MAHNOOR)
Our new approach to mentor is evolving into a Multimodal AI Learning assistant whose main objective will be providing intentionally focused information through feedback loops in order to develop the necessary knowledge base and critical thinking skills that will allow the development of higher levels of proficiency in any educational or professional setting.

Slide 19 - Build an AI… (MAHNOOR)
We'll begin by developing an AI powered cognitive tool, that will transition between modes (text, images and spatial modeling), to then engage it in a collaborative dynamics with the user with explainable AI feedback pipeline, after we can run test scenarios to compare models in order to refine outputs to finally be able to measure how users respond and learn from such feedback.

Slide 20 - Building a Cognitive tool (MAHNOOR)
In that sense, Mentor learning assistant will be engineered as a multimodal cognitive tool capable of analyzing contextual needs, delivering focused information strategically while engaging critical thinking processes and creating feedback loops for growth for building a sustainable proficiency in any subject it is trained for.

Slide 21 - A Reliable Path (MAHNOOR)
Since we're going to be focused in a very narrow area of the Architectural knowledge base for our proof of concept, we'll be staging it in three distinctive phases: Ideation, Visualization and Materialization.

Slide 22 - 3 Phases explained (MAHNOOR)
This pipeline outlines an AI-augmented architectural feedback system structured across three phases: Ideation, Visualization, and Materialization. The user begins with a cognitive warm-up and prompt scaffold, preparing input for the system. In the Visualization phase, an uploaded image undergoes computer vision preprocessing, region segmentation, and vision-language analysis, leading to mapped critique anchors and persona-tuned feedback. The Materialization phase expands this to 3D models, parsing geometry, performing spatial analysis, and generating semantic labels and critiques via LLMs, which are overlaid in a 3D interface. Each phase feeds into a feedback loop for prompt refinement and user learning, all connected through a dynamic data layer.

Slide 23 - A combined strategy (MAHNOOR)
And in order to deliver each phase as truthfully to our mission as possible, we'll be putting in place 4 key protocols throughout the processes:
-**Context Reasoning** in order to interpret and solve problems by considering the surrounding circumstances, background information, and relevant details of a scenario.
-**Knowledge Synthesis** to systematically gather, appraise, synthesize, and analyze existing research to provide a comprehensive understanding of a specific topic.
-**Socratic Dialogue** as an inquiry method to use a question-and-answer format to explore underlying beliefs and assumptions, ultimately leading to a deeper understanding of a topic.
-**And additional practice area specific** methods to track the metacognitive development of the

learning process.

Slide 24 - 4 Agentic Protocols explained (MAHNOOR)
Feeding into one another, our four agents work hand in hand.
The first Agent analyzes space using computer vision and 3D processing. It provides visual and data-driven outputs, but the user must interpret their spatial significance. "Here’s the visibility analysis, what spatial qualities does it suggest?"
The second Agent activates when the user makes spatial observations. It offers relevant information, principles, and precedents, but still relies on user interpretation. "Your proportion strategy relates to these historical design principles, how do they apply to your context?"
The third Agent presents alternative perspectives and constructive challenges, prompting critical reflection. "This design choice could lead to certain consequences, how might you address them?"
The fourth Agent tracks the user's reasoning over time, offering meta-level feedback. "You consistently excel in circulation analysis but often overlook lighting relationships. Here’s the data, shall we focus there next?"

Slide 25 - Questions (MAHNOOR)
Does this mean we have it all figured out already? Far from it! Here are just a few of the questions that have arisen along the way.

Slide 26 - Q1 Capability (MAHNOOR)
Can multimodal AI systems trained on comprehensive architectural knowledge analyze, extract and interpret architectural spatial relationships from 2D plans and 3D models effectively and integrate spatial analysis with theoretical frameworks to generate expert-level design feedback?

Slide 27 - Q2 Collective Intelligence (MAHNOOR)
What emergent insights and solutions arise from collective multimodal intelligence that are impossible to achieve through individual multimodal agents or single-modal systems?

Slide 28 - Q3 Interactivity (MAHNOOR)
How does interactive multimodal learning that combines visual spatial analysis with conversational development positively affect cognitive enhancement and skill transfer and at what point it begins to reduce designer agency?

Slide 29 - Q4 Integration (MAHNOOR)
How effectively can multimodal AI systems integrate into professional spatial design workflows while maintaining quality standards and enhancing rather than replacing human expertise?

Slide 30 - SOTA (SEDA)
Hopefully we don’t have to do this alone, with your help and some of the State of the Art research currently underway, we hope to eventually give answers to some of those questions.

Slide 31 - Tools Oversight (SEDA)

We begin with **SketchUp AI Critic**, a vision-language plugin using GPT-4o to provide design feedback directly inside SketchUp.
GODESIGN and **HSC-GPT** push this further—GODESIGN offers stylistically grounded critique based on architectural philosophy, while HSC-GPT is a domain-specific large language model trained for human settlement and planning contexts.
Autogen Studio supports multi-agent design critique setups, enabling parallel feedback streams.
On the spatial intelligence front, **TestFit** and **Hypar** facilitate early-stage massing and layout optimization, with Hypar allowing script-based generation of IFC-compliant 3D models.
Finally, **Visual Prompting Critique** and **Socratic Models** represent frameworks for multimodal AI dialogue and critique logic—tools that deeply influence how we structure AI feedback loops in design contexts.

Slide 32 - Categories Description (SEDA)
Here, we distill the four main domains that shape our thesis:
Multimodal AI provides the technical backbone—bridging vision, text, and spatial reasoning.
Architecture + AI explores how computational models interact with built environment logic and design processes.
Human + AI examines the interface between designer cognition and machine augmentation—particularly relevant for real-time critique and cognitive scaffolding.
And **Design Feedback** centers our inquiry on the communicative, iterative, and evaluative nature of critique, drawing from studio culture and computational design practices.
These four pillars allow us to connect tools, methods, and user experience into one coherent architectural intelligence pipeline.

Slide 33 - Models Overview (SEDA)
This slide maps a broader landscape of related research and tools.
In the **Multimodal AI** category, we find tools like **DreamSketch**, **Text2Scene**, and **MM-REACT**, all of which turn sketches or scenes into richly described outputs.
Under **Architecture + AI**, tools like **Hypar**, **AI-Assisted Design**, and **Project Discover** showcase performance-informed generation and layout optimization.
The **Human + AI** section includes collaborative models like **Creative Colleague**, **Co-creator**, and **UX Interpreter**, reflecting co-piloting frameworks.
Lastly, **Design Feedback** gathers tools like **ArchiJury** and cognitive prompting frameworks like **Chain of Thought**, which influence how AI provides structured, reflective critique.

Slide 34 - Multimodal AI (SEDA)
Here is where we introduce Multimodal AI examples

Slide 35 - Architecture + AI (SEDA)
Here is where we introduce Architecture + AI examples

Slide 36 - Human + AI (SEDA)
Here is where we introduce Human + AI examples

Slide 37 - Design Feedback (SEDA)
Here is where we introduce Design Feedback examples

Slide 38 - Resources and reference links (SEDA)
Here is our growing list of references and resources in case anyone would like to dive deeper in each one of them.

Slide 39 - Timeline (SEDA)
And this is how we're strategizing our work in the limited amount of time we have available for this

Slide 40 - Timeline Phasing Introduction (SEDA)
Here is where we explain the different phases of development

Slide 41 - Schedule (SEDA)
Here is where we explain the schedule

Thesis Review 2 Script

Slide 42 - Testing (BIEL)
So one of the things we had to start working with were the testing methods for Mentor

Slide 43 - Participant Selection & Grouping (BIEL)
Beyond the initial internal testing that Mahnoor, Seda and Myself will conduct while refining Mentor, we've thought that we will require a much larger test group for mentor, the conventional AI tools with which it will be benchmark against and the control group. Our selection process intentionally will target all groups with similar educational background, for which we're planning on engaging the rest of the Master's program cohort to test Mentor (if time and participation allows for it). They'll be equally divided between test groups for mentor and other conventional tools and a control group, and they will all be assigned the same design tasks for measurable consistency.

Slide 44 - Sample Size Overview (BIEL)
This means that our ideal testing scenario will count with the rest of the class, close to 30 students in total (including us), the group selection will be randomized, with equal partition between experimental and control conditions in order to ensure comparable baseline conditions.

Slide 45 - Task Design (BIEL)
We will create a set of architectural design challenge scenarios in which both creativity and critical thinking are required, rather than tasks with a single correct answer.

Slide 46 - Task Design (BIEL)
Each participant will undertake several tasks of varying complexity over the study period, which will be structured in logical progression phases such as the ideation phase, the visualization phase and the materialization phase, each one requiring specifying critical thinking and problem solving skills in order to be completed before proceeding to the next stage.

Slide 47 - Task Structure (BIEL)
Each challenge will follow the same structural timeline by phases, each of which will be equally split into a dual assessment framework. One which will focus on the quality of the process while the other will focus on the quality of the output, facilitating our capacity to measure each one of those KPI's along the way.

Slide 48 - Data Collection (BIEL)
But what data we'll be exactly collecting (and how?)

Slide 49 - Data Collection (BIEL)
We'll be focusing our efforts in a combined strategy that will combine the multi agent interaction, the multimodal visual analysis but hopefully also a biometric component through an eye tracking system.

Slide 50 - Data Collection (BIEL)
Our primary framework will be our multiagentic scaffolding system, which will be the one in charge of enforcing the critical thinking and problem solving skills we are trying to improve with mentor, while our multimodal analysis of visual content will help us extract a completely different set of features based not only on the analysis of the visual inputs, but also the generation of visual outputs along the way. Lastly, and in order to tie up our analysis with those performed during the MIT tests, we wanted to incorporate at least a superficial level of biometric feedback, which in our case will translate into an eye tracking solution which will help us measure focus, engagement and attention patterns and cognitive load via pupil dilation and gaze-content alignment.

Slide 51 - Data Synchronization Architecture (BIEL)
Then, a Lab Streaming Layer (LSL) will serve as the central nervous system for our multi-agentic data collection framework, it will elegantly orchestrate the dialogue and visual logs captured via OpenTelemetry standards, high-resolution screen recordings, eye tracking data streams, and further behavioral observations.

The synchronized data will flow into a cloud-based data lake using XDF format, ensuring that every interaction, gaze pattern, and physiological response is precisely time-aligned for comprehensive multimodal analysis.

Slide 52 - Practical Implementation (BIEL)

This measurement framework operationalizes the AI-human interaction through four key metrics:
1.- Acceptance rates measuring verbatim AI suggestion usage
2.- Modification indices calculated via edit distance algorithms
3.- Time-to-decision latencies capturing cognitive processing
4.- and usage pattern analysis revealing help-seeking behaviors.

While the visual content processing will follow a systematic pipeline from capture through feature extraction to a classification system, ultimately correlating visual output quality with learning outcomes. through data-driven pattern recognition.

Slide 53 - Performance (BIEL)
After completing the tasks, we will measure performance on multiple levels:

First by assessing the quality of their solution to the task, which we'll score on criteria such as creativity, functionality and reasoning quality.

Then we'll process and analyze the logs transcripts based on critical thinking engagement, we'll analyze if the student reflection was autonomous of AI aided and will compare the time taken during a number of interactions.

To then score the cognitive measures during pre, during and post task performance.

To finally culminate with an added layer of information through a user feedback survey in which students will self-evaluate their level of engagement during the different tasks and tools used per group.

Slide 54 - Benchmarking (BIEL)
Now after defining the testing methods, the tasks and the kind of data we'll be collecting and how, is time to define how we'll be benchmarking our model.

Slide 55 - Benchmarking Objectives (BIEL)
Our benchmarking objectives will include both standard performance metrics and novel metrics for cognitive outcomes.

In essence we'll simulate authentic learning scenarios while maintaining control for scientific analysis. By comparing the mentor against other conditions on the same tasks, we can pinpoint its effects on both performance and cognitive growth.

Slide 56 - Established LLM Benchmarks (BIEL)
Looking at some already established LLM Benchmarking methods we found the one from the American Invitational Mathematics Exam, which is a challenging math test from high school competitions. HellaSwag, a commonsense reasoning task where the model must pick the

sensible continuation of a story from plausible alternatives. MT-Bench, which measures a multi-turn conversational and reasoning ability (it emphasizes updating understanding over a dialogue) and TruthfulQA which tests factual correctness and resistance to false premises (it labels each answer as correct or hallucinated)

Slide 57 - Open-Source Benchmarking Framework **(BIEL)**
Alternatively, we can also adopt and open-source benchmarking framework like Livebench, which is basically A toolkit with diverse tasks that avoids test-set contamination by releasing new questions monthly

Slide 58 - Generalized AI Agent Benchmark Framework (GAIA) **(BIEL)**
Another path for effective benchmarking could be a Generalized AI Agent Benchmark Framework in which we assess our agent's ability to operate under diverse conditions, adapt to changing scenarios, collaborate & generalize knowledge.

Slide 59 - Phasing Implementation Strategy **(BIEL)**
Regardless of whichever route we end up choosing as our path for effective benchmarking, industry standards agree that the best strategy for implementation should consider 4 kay phases for development and implementation.

Slide 60 - Foundation and Framework Setup **(BIEL)**
The first of those phases will be the foundation and framework setup, in which we have few options to establish our baseline for Cognitive Assessment Infrastructure, but also our competitive model configuration for target models and define our task standardization.

Slide 61 - Pilot Testing & Calibration **(BIEL)**
During our pilot testing and calibration phase, we'll conduct our controlled pilot study (within the 3 of us as subjects), will collect and fine-tune our first Key Performance indicators and refine our data collection protocols for each one of the methods and test-groups.

Slide 62 - Full-Scale Comparative Analysis **(BIEL)**
Our most challenging phase because of the recruiting demands and external variables (not to mention timing constraints) will be the Full-Scale Comparative Analysis phase, in which we'll expand the group to the whole student cohort and incorporate advanced comparative metrics, while we try to also project a cost analysis for potential scalability beyond this group.

Slide 63 - Analysis and validation **(BIEL)**
Finally, during our Analysis and validation phase we'll focus on the statistical analysis framework of the results obtained while establishing the validation protocols for Mentor.

Slide 64 - Benchmarking evaluation **(BIEL)**
As a result, we expect to obtain a comprehensive benchmarking evaluation framework between Mentor and other available models with insights into the cognitive development differences

between them, performance differences, interaction patterns, performance profiles and a comprehensive outline of the risks of incurring into cognitive offloading on each one of them.

Slide 65-74 - Questions (Mahnoor)

Slide 75-80- Agents (SEDA)

Let`s introduce our agents. The core principle is simple: we never give students direct answers. Instead, we use four specialized AI agents that work together to guide students toward discovering solutions themselves. Think of it like having four different types of teachers in one system - each with their own expertise, but all coordinating to help the student learn. The system runs on LangGraph, which acts as our orchestrator. When a student types something, the system first analyzes what they need, then decides which agents should respond. Sometimes it's just one agent, sometimes all four work together. The key is that everything is designed to make students think, not just absorb information.

Core Components

Four specialized agents with distinct educational roles

LangGraph orchestration for agent coordination

Shared memory system tracking student progress

Domain-flexible architecture supporting multiple subjects

Decision Logic

Input analysis determines student needs

Router selects appropriate agent combination

Agents process using specialized knowledge and methods

System synthesizes responses prioritizing active learning

Slide 81-82 Context Reasoning (SEDA)

Before any teaching happens, we need to understand what the student actually needs.Our Context Reasoning Agent is here to read between the lines. When a student sends a message, this agent analyzes not just the words, but the emotions, the confidence level, how engaged they seem. It looks at things like: Are they confused? Overconfident? Asking a simple question or struggling with a complex concept? Based on this analysis, it creates what we call a 'context package' - basically a detailed profile of what this student needs right now.

Analysis Functions

Linguistic processing to understand complexity level

Emotional state detection for confidence and engagement

Understanding assessment through vocabulary analysis

Learning context classification for appropriate response type

Output Classifications

Student understanding: low, medium, or high comprehension

Confidence levels: uncertain, confident, or overconfident

Engagement metrics: passive, active, or highly engaged

Interaction type: factual query, exploration, or design feedback

Slide 83 - Vision Pipeline (Mahnoor)
The workflow starts with sketches or uploads, translating images using models like YOLO and SAM. Inputs are structured via bounding boxes, masks, and semantic overlays, moving seamlessly from extraction to smart analysis that links data-driven insight with interactive critique for creative advancement.

Slide 84 – Multi-Agent Tools & Orchestration (Mahnoor)

Multiple tools—OpenCV, YOLO, Claude, GPT, and Magma—collaborate in an orchestrated system. Supervisors like CrewAI and LangGraph help synthesize multimodal data, enabling nuanced critique and promoting user growth and metacognition.

Slide 85 – Detection (Mahnoor)

Segmentation uses SAM and YOLO for pixel-level masks and zoning. We trained a custom YOLO (You Only Look Once) model specifically for architectural element detection using a dataset of 103 architectural classes including structural elements (walls, doors, windows), functional spaces (bedrooms, kitchens, bathrooms), and design elements (corridors, circulation nodes). The training dataset consisted of annotated floor plans and elevations with augmented variations to improve model robustness.
This is, however, not accurate and needs a lot more work.

Slide 86 - Segmentation (Mahnoor)

This slide outlines the segmentation workflow using the Segment Anything Model (SAM), specifically the ViT-H variant with 4 billion parameters and a prompt encoder. The process begins by either using YOLO for object detection or proceeding without it. When YOLO is used, bounding boxes guide the mask generation or point-based detection prompts the segmentation. If YOLO is unavailable, the system falls back on CLIP or BLIP for image captioning and uses the generated text to guide segmentation. This dual-path approach ensures flexible and robust object segmentation under varying input conditions.

Slide 87 - Semantic Understanding (Mahnoor)

This slide explains the *Semantic Understanding* component of the workflow, which leverages CLIP and BLIP for evaluating architectural designs through natural language. CLIP matches RGB images with text prompts to generate similarity scores, while BLIP generates descriptive captions for the same images. These outputs are analyzed to detect design issues—such as limited circulation or poor window placement—by checking if specific concept scores fall below a threshold. The system then translates these insights into design metrics like circulation efficiency and lighting quality, which feed into critique points. Despite producing meaningful feedback, limitations include imprecise object detection and the fact that these models aren't originally trained on architectural design criteria. The reported accuracy stands at 78% for CLIP and 82% for BLIP.

Slide 88 - Pipeline (Mahnoor)

This slide presents the full pipeline for the system, detailing each stage from image preprocessing to final output. It begins with OpenCV enhancing raw images, followed by YOLOv8 detecting objects like walls and windows. SAM then segments these detections into pixel-level masks. CLIP and BLIP provide semantic understanding—CLIP assigns concept scores based on image-text alignment, while BLIP generates descriptive captions. A custom module analyzes all outputs to calculate design metrics such as circulation and lighting scores. Another module critiques the design with feedback like “Add more windows,” and the final stage outputs annotated images and a report.

Slide 89 - Training (Mahnoor)

a comprehensive framework for evaluating architectural design, divided into four main categories: Architectural Elements, Analysis Capability, Critique Categories, and Design Metrics. Each section lists critical considerations such as structural and spatial elements, types of analytical evaluation (spatial, functional, technical, aesthetic), areas for critique (like functionality, aesthetics, technical issues), and measurable performance metrics (such as circulation efficiency and accessibility). Together, these categories provide a structured approach to assessing and improving architectural projects, ensuring coverage of both qualitative and quantitative aspects.

Slides 90 - 95 - Training Progress (Mahnoor)

Showing how we tested for a custom YOLO model using our own dataset to see if it would respond to the training and give better image analyses. While it did improve, the fact that the dataset was not very large might have impeded on how much it could learn from and then apply to the sample images it was provided.

Slide 96-97 - Knowledge Synthesis Agent (SEDA)

This one is like a search engine that never just hands the information - it helps you find the right information and asks questions to make sure you understand it.

The agent uses vector databases - Vector embedding converts student queries to searchable format, ChromaDB collections store domain-specific knowledge bases, Similarity scoring determines content relevance and quality, if the information is not available in the database then it looks for predefined websites to find relevant information and gives the source link back to user to do search for further if they want. In the end Adaptive synthesis matches complexity to student comprehension

This Response Strategy is if no relevant content found: Generate exploratory questions or do web search, Medium-quality matches: Use as supporting evidence with clarification, High-quality matches: Synthesize information while maintaining inquiry

Slide 98-100 - Socratic Dialogue Agent (SEDA)

Socratic Dialogue Agent

Our Socratic Agent embodies the ancient art of learning through questions - like Socrates did, but powered by AI. But it's not random questioning - it's strategic. If a student seems confused, it

asks clarifying questions to help them organize their thoughts. If they're confident, it pushes deeper with analytical questions. If they're overconfident, it challenges their assumptions. The goal is always the same: guide the student to that 'aha!' moment where they discover the answer themselves.Question Strategy Framework is
Understanding-based adaptation: Clarification for beginners, analysis for advanced
Confidence calibration: Support for uncertain, challenges for overconfident
Progressive complexity building from simple to sophisticated inquiry
Five question types: exploration, clarification, analysis, synthesis, evaluation
Implementation Principles
Never provide direct answers under any circumstances
Build on student's existing vocabulary and concepts
Escalate complexity gradually based on demonstrated understanding
Encourage verbalization of reasoning processes

Slide 101-102 - Cognitive Enhancement Agent (SEDA)
This agent is like a personal trainer for the mind. It constantly monitors how hard the student's brain is working and adjusts the challenge level accordingly.If a student seems bored or disengaged, it throws them a curveball - maybe asking 'What if your budget was cut in half?' or 'How would a child experience this space?' If they seem overwhelmed, it helps break down the problem. If they're in that sweet spot where they're challenged but not frustrated, it maintains that level. The goal is to keep students in what we call the 'zone of proximal development' - challenged enough to grow, but not so much that they give up."
Monitoring Systems:Real-time engagement tracking through response patterns, Cognitive load assessment via confusion and processing indicators, Metacognitive awareness measurement through self-reflection quality, Behavioral consistency analysis for learning pattern identification
Intervention Techniques
Challenge scenarios: Constraint modifications to stimulate problem-solving
Perspective shifts: Alternative viewpoints to broaden thinking
Metacognitive prompts: Questions about thinking processes themselves
Complexity adjustments: Dynamic difficulty scaling based on performance

Slide 103-104- Tracking (SEDA)
When a student sends a message, the Context Agent analyzes it, the Router decides which agents to activate, and they all work together to create a response. But here's the key - we prioritize Socratic questions over everything else, then cognitive challenges, then information. We never lead with facts. The system learns constantly. It tracks what works for each student, adapts across different subjects, and improves its teaching over time. Whether we're teaching architecture or game design, the core approach stays the same, but the knowledge and examples change completely. The result is an AI tutor that truly guides learning rather than just delivering information.
Integration Process is Multi-agent coordination through priority-based response synthesis,Quality assurance ensuring educational value and appropriate challenge,Cross-session learning with student profile development,Domain flexibility enabling seamless subject transitions

Success Metrics
Student engagement measured through participation depth and quality
Learning progression tracked via concept mastery and skill development
System performance evaluated through response relevance and coordination
Adaptive capability assessed through cross-domain effectiveness

Slide 105-106- Workflow (SEDA)