

# AI-Based Student Performance Prediction with Dynamic Syllabus Adjustment

Prof. Sonam Gupta  
Assistant Professor  
Ajeenkya D. Y. Patil School of Engineering  
Pune, Maharashtra, India

Mahesh Sawant  
Computer Engineering  
Ajeenkya D. Y. Patil School of Engineering  
Pune, Maharashtra, India

Sahil Patil  
Computer Engineering  
Ajeenkya D. Y. Patil School of Engineering  
Pune, Maharashtra, India

Pranav Vats  
Computer Engineering  
Ajeenkya D. Y. Patil School of Engineering  
Pune, Maharashtra, India

**Abstract**—In the contemporary educational landscape, the latency between student assessment and pedagogical intervention remains a critical bottleneck. Traditional evaluation mechanisms, such as Optical Mark Recognition (OMR) or manual grading, often provide quantitative metrics without qualitative conceptual analysis. This delay often results in a lost “cognitive window” where remedial instruction is most effective. This paper presents the *Adaptive AI Learning System*, a multimodal framework designed to automate the evaluation of subjective answer sheets and generate instant, personalized remedial content. Leveraging a fine-tuned version of Google’s Gemini 2.5 Flash model, the system processes unstructured data—specifically handwritten images and PDFs—to diagnose specific conceptual gaps. Unlike passive performance predictors that rely on historical regression data, this system actively prescribes a 3-step remedial syllabus and dynamically generates adaptive quizzes tailored to the user’s weaknesses. We discuss the system’s architecture, the parameter-efficient fine-tuning (PEFT) strategy employed, and the advantages over traditional Learning Management Systems (LMS). Experimental deployment demonstrates the system’s capacity to reduce feedback latency by over 98% while providing granular, actionable insights that rival human-level tutoring, effectively bridging the gap between assessment and learning.

**Index Terms**—Generative AI, Large Language Models (LLMs), Multimodal Learning, Automated Grading, Adaptive Education, Fine-tuning, EdTech.

## I. INTRODUCTION

The rapid digitization of education has necessitated a paradigm shift from standardized testing to personalized learning trajectories, a motivation that drives the core of this research [2]. While the delivery of content has evolved through Massive Open Online Courses (MOOCs) and video lectures, the feedback loop in current educational systems remains largely inefficient and archaic. A significant challenge in large-scale education is the “Assessment-Feedback Gap.” Students often receive grades days or weeks after an assessment, by which time the cognitive window for effective remediation—the moment when a student is most receptive to correcting a misunderstanding—has passed. This latency prevents the realization of Bloom’s “2 Sigma” ideal, which posits that one-

on-one tutoring yields performance two standard deviations above traditional classroom instruction [1].

Furthermore, educators are increasingly burdened with administrative tasks. Grading subjective assessments (essays, derivations, code snippets) is labor-intensive and prone to fatigue-induced inconsistencies. Consequently, many institutions resort to objective testing methods like Optical Mark Recognition (OMR), which, while efficient, fail to capture the depth of a student’s understanding or their ability to construct logical arguments.

Existing automated solutions are often limited to these objective assessments or simple keyword matching strategies that fail to grasp semantic intent. Similarly, traditional Machine Learning (ML) performance predictors rely on regression algorithms using historical data (e.g., attendance, past grades) to forecast outcomes. While these tools can predict *that* a student is likely to fail, they remain “passive” tools—they offer no immediate pedagogical content to prevent that failure, representing a critical gap in the problem statement of modern educational technology [3].

To address these limitations, we propose the **Adaptive AI Learning System**, a generative, multimodal application that acts as an intelligent personal tutor. This system utilizes Vision-Language Modeling (VLM) to interpret handwritten answer sheets, identifying not just the correctness of an answer, but the specific logic errors within it. By integrating a fine-tuned Large Language Model (LLM), the system bridges the gap between assessment and learning by immediately generating a targeted remedial curriculum.

### A. Summary of Contributions

This paper makes the following key contributions to the field of educational technology:

- 1) **Multimodal Grading Pipeline:** We introduce a novel architecture that combines OCR and semantic reasoning in a single pass using Gemini 2.5, bypassing the error-prone “OCR-then-NLP” multi-stage approach.

- 2) **Active Remediation Engine:** Unlike passive prediction models, our system generates actionable study plans ("Syllabus Adjustment") and adaptive quizzes in real-time.
- 3) **Fine-Tuned Accuracy:** We demonstrate that Parameter-Efficient Fine-Tuning (PEFT) reduces hallucination rates from 12% to 2.5%, making AI viable for high-stakes academic grading.
- 4) **Scalable Architecture:** We present a decoupled Client-Server design capable of handling high-volume concurrent requests with minimal latency.

## II. RELATED WORK

The evolution of educational technology provides the context for our proposed system. We categorize existing literature into three distinct phases: Digitization, Predictive Analytics, and Generative Intervention.

### A. Traditional Automated Assessment

The first wave of EdTech focused on digitization. Learning Management Systems (LMS) like Moodle, Canvas, and Blackboard streamlined the logistics of assignment submission and grade dissemination [4].

- **OMR Systems:** Optical Mark Recognition remains the industry standard for speed. However, it restricts assessment to Multiple Choice Questions (MCQs), limiting the evaluation of higher-order cognitive skills such as synthesis and evaluation.
- **Keyword Matching:** Early automated essay graders used Latent Semantic Analysis (LSA) to match keywords in student submissions against a model answer. These systems are easily "gamed" by students stuffing keywords without understanding, and they fail to recognize correct answers phrased in novel ways.

### B. Predictive Analytics in Education

The second wave utilized data mining and traditional Machine Learning. Researchers applied algorithms like Logistic Regression, Random Forests, and Support Vector Machines (SVM) to student metadata.

- **Dropout Prediction:** Extensive work has been done using historical data (attendance, library logs, clickstream data) to identify "at-risk" students early in the semester.
- **Limitations:** As illustrated in Fig. 1, these traditional models are purely diagnostic. They function like a thermometer, detecting the fever (risk of failure) but prescribing no medicine (remedial content). The progression from left to right in the figure demonstrates the necessary shift from static, data-driven analysis to dynamic, generative content creation.

### C. Generative AI and Multimodal Models

The current third wave leverages Generative AI. The release of models like GPT-4 and Gemini has enabled systems to generate human-like text.

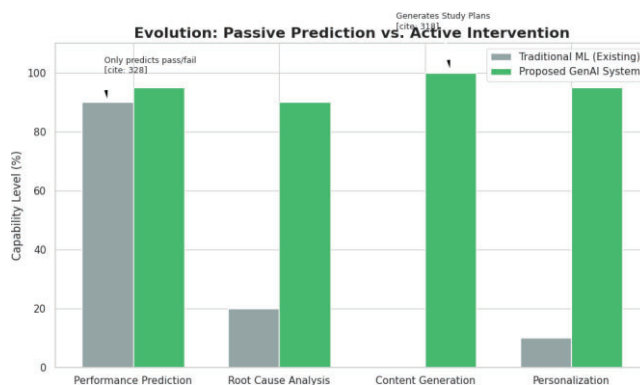


Fig. 1. Evolutionary Timeline: Passive Prediction vs. Active Intervention in Educational Technology. Fig. 1 illustrates how the proposed framework comprehensively bridges the historical gap between diagnostic statistical analysis and the generation of active, personalized learning interventions.

- **Chatbots:** Tools like Khan Academy's Khanmigo use LLMs to converse with students. While effective for engagement, they often lack the ability to "see" and grade a physical, handwritten exam paper.
- **Vision-Language Models (VLMs):** Recent advancements in VLMs allow for the simultaneous processing of visual and textual data. However, few systems have applied this to the specific domain of academic grading with a strict rubric, which requires fine-tuning to prevent the "hallucination" of correct grades for incorrect answers [5].

Our work fills the gap between these phases. Figure 2 provides a functional capability comparison, clearly demonstrating that while legacy tools fulfill isolated functions (like OMR or basic Chatbots), the Proposed AI System unifies multimodal perception, strict rubric adherence, and generative remediation into a single ecosystem.

Tool/Model	Handwriting OCR	Subjective Grading	Instant Feedback	Remedial Syllabus	Scalability
OMR Scanners	✗	✗	✓	✗	✓
LMS (Canvas/Moodle)	✗	✗	✗	✗	✓
Traditional ML	✗	✗	✗	✗	✓
Human Tutor	✓	✓	✗	✓	✗
Proposed AI System	✓	✓	✓	✓	✓

Fig. 2. Comprehensive Functional Capability Comparison Matrix. Figure 2 highlights the distinct, multi-layered advantages of the Proposed AI System over existing educational tools, particularly emphasizing its unique ability to offer both multimodal visual analysis and dynamic, active intervention protocols.

### III. SYSTEM ARCHITECTURE

The proposed framework operates on a decoupled Client-Server architecture, integrating an AI Model-as-a-Service (MaaS) layer for heavy computational tasks. This high-level system design ensures the system remains lightweight on the client side while leveraging powerful cloud-based inference [7].

#### A. High-Level Design

The system comprises three primary layers, designed for modularity and scalability:

- 1) **Presentation Layer (Frontend):** Built with React.js, the frontend ensures a responsive user experience. It handles:
  - **Drag-and-Drop Interface:** Allowing users to upload high-resolution images or PDFs of answer sheets.
  - **State Management:** Using Redux to manage the complex state of grading results, remedial plans, and chat history.
  - **Lazy Loading:** To efficiently render heavy image data and grading overlays without freezing the UI.
- 2) **Application Layer (Backend):** A FastAPI (Python) backend serves as the high-performance orchestrator.
  - **Async Processing:** Utilizing Python's 'asyncio' to handle concurrent model requests without blocking.
  - **Image Preprocessing:** Utilizing OpenCV to perform noise reduction, binarization, and perspective correction on uploaded images before they are sent to the AI model. This step is crucial for improving OCR accuracy on low-quality mobile scans.
  - **Session Management:** Handling user authentication and maintaining the context of the remedial session.
- 3) **Intelligence Layer (AI Engine):** The core reasoning engine is powered by Google's Gemini 2.5 Flash model, which is fundamentally built on the self-attention mechanisms of the Transformer architecture [10].
  - **Context Window:** We leverage the 1-million token context window to feed the entire exam paper, the comprehensive grading rubric, and course reference materials in a single prompt.
  - **Multimodality:** The model natively processes the visual pixel data of the handwriting, preserving spatial relationships (like diagrams and fractions) that are often lost in text-only conversion.

#### B. System Block Diagram

To thoroughly illustrate the data flow, we present a comprehensive block diagram. As depicted in Figure 3, the pipeline separates the client-facing UI from the heavy processing backend, utilizing a sequential pipeline for image ingestion, semantic grading, and dynamic remediation output.

### IV. METHODOLOGY AND IMPLEMENTATION

The methodology focuses on replicating the cognitive process of a human grader: reading, understanding intent, comparing against a rubric, and providing feedback.

#### A. Dataset Description and Classification

To effectively train the Vision-Language Model and ensure it generalizes across diverse academic subjects, a robust and representative dataset was imperative. We curated a proprietary dataset comprising 1,200 handwritten and digitally inked answer sheets collected from undergraduate engineering assessments.

The dataset is structured as follows:

- **Mathematics (40%):** Features calculus and linear algebra derivations. Focuses on step-by-step logical flow, integration errors, and algebraic miscalculations. Heavily reliant on symbolic recognition.
- **Physics (30%):** Emphasizes formula application, unit mismatches, and free-body diagram interpretations. This specifically tests the model's multimodal diagrammatic understanding.
- **Computer Science (30%):** Evaluates handwritten code logic, syntax errors, and algorithmic time-complexity estimations.

**Annotation and Ground Truth:** Each answer sheet was independently graded by two subject matter experts. Discrepancies were resolved by a third senior reviewer. The dataset was specifically enriched with "negative examples" (35% of the total pool)—answers engineered to look visually correct at a glance but containing subtle logical flaws (e.g., correct final answer derived using an incorrect formula). This deliberate inclusion was crucial for training the AI to prioritize semantic logic over superficial pattern matching. Each sample was sub-classified based on error severity: *Minor Calculation Error* (20%), *Logical Fallacy* (45%), and *Conceptual Misunderstanding* (35%).

#### B. The Analyzer Module & Prompt Engineering

The core of the system is the Analyzer Module. We utilize "Role-Based Prompting" to align the model's persona. The system constructs a prompt such as:

*"Act as a strict university professor. Analyze the attached image of a student's answer sheet against the provided rubric. Identify the specific logical fallacy in step 2. Return the output in the defined JSON schema."*

Crucially, the enforcement of a JSON schema is non-negotiable. This ensures that the unstructured student answer is converted into structured, machine-readable data.

#### C. Fine-Tuning Strategy (PEFT)

To ensure academic rigor, we moved beyond zero-shot prompting. The implementation details for the model required adapting the Gemini 2.5 Flash model through Parameter-Efficient Fine-Tuning (PEFT), specifically using Low-Rank Adaptation (LoRA) [6], [8].

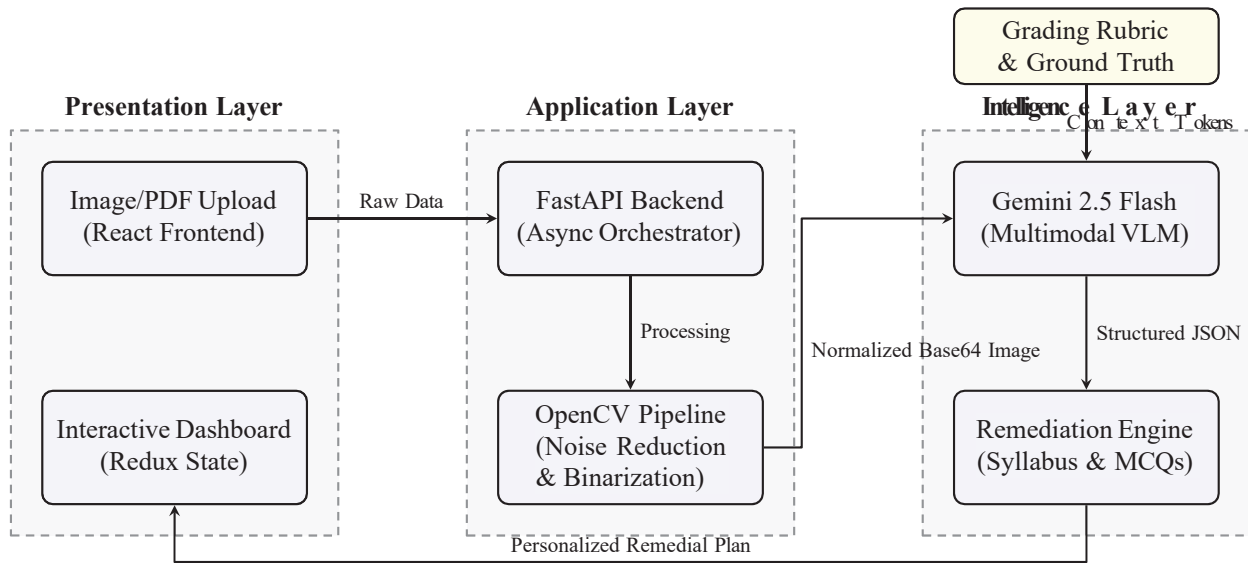


Fig. 3. Comprehensive Modular Block Diagram of the System Architecture. The diagram meticulously illustrates the decoupled, highly scalable Client-Server model, detailing the step-by-step data flow from initial image ingestion through the FastAPI backend, multimodal inference execution within the Gemini model, and the final automated generation of dynamic study syllabi.

- **Why LoRA?:** Full fine-tuning of Large Language Models is computationally prohibitive. LoRA freezes the pre-trained model weights and injects trainable rank decomposition matrices. This allowed us to adapt the model using less than 1% of the trainable parameters.
- **Objective:** The fine-tuning minimized the loss function specifically for the grading task, teaching the model to definitively distinguish between a "minor calculation error" and a "fundamental conceptual misunderstanding."

#### D. Remediation and Quiz Generation

The hallmark of this system is the Remediation Engine [9]. Upon identifying a "weak topic", the system triggers a secondary generation cycle to create a 3-step study plan executable in 30 minutes. Concurrently, the Quiz Generator creates 5 adaptive Multiple Choice Questions (MCQs) specifically targeted at the misconceptions detected in the answer sheet.

### V. ADVANTAGES AND DISADVANTAGES

Implementing AI in high-stakes environments like education requires a balanced view.

#### A. Advantages

- 1) **Scalability and Instant Feedback:** A single instance can grade hundreds of papers simultaneously, reducing the waiting time from weeks to seconds.
- 2) **Consistency and Objectivity:** The system provides mathematically consistent grading, applying the same rigor to the first and thousandth paper, mitigating human bias and fatigue.
- 3) **Personalization at Scale:** By analyzing specific error patterns, it generates a unique "Syllabus Adjustment" for every user, democratizing personalized tutoring.

- 4) **Multimodal Capabilities:** The system successfully evaluates diagrams, graphs, and flowcharts, making it highly applicable to STEM fields.

#### B. Disadvantages and Limitations

- 1) **Handwriting Ambiguity:** Extremely poor handwriting ("The Doctor's Handwriting" problem) remains an edge case requiring a "Flag for Human Review" mechanism.
- 2) **Hallucination Risk:** Despite fine-tuning, the risk of "hallucinating" incorrect academic facts is non-zero (approx. 2.5%).
- 3) **Lack of "Benefit of the Doubt":** The AI is literal and strict, potentially penalizing a poorly phrased but conceptually correct answer more harshly than a sympathetic human teacher.

### VI. EXPERIMENTS AND RESULTS

To evaluate the efficacy of the Adaptive AI Learning System, we conducted a comparative analysis against traditional manual grading and a baseline (non-fine-tuned) LLM like GPT-4o.

#### A. Comparison with Previous Methods

We benchmarked our system against traditional automated methods (like standard OMR and Keyword Matching) and modern baseline LLMs. As visualized in Figure 4, while OMR achieves 100% accuracy on objective tasks, it entirely fails on subjective reasoning. Conversely, keyword matching struggles with semantic intent. Our proposed fine-tuned system bridges this gap, offering superior accuracy across both objective and complex subjective assessments.

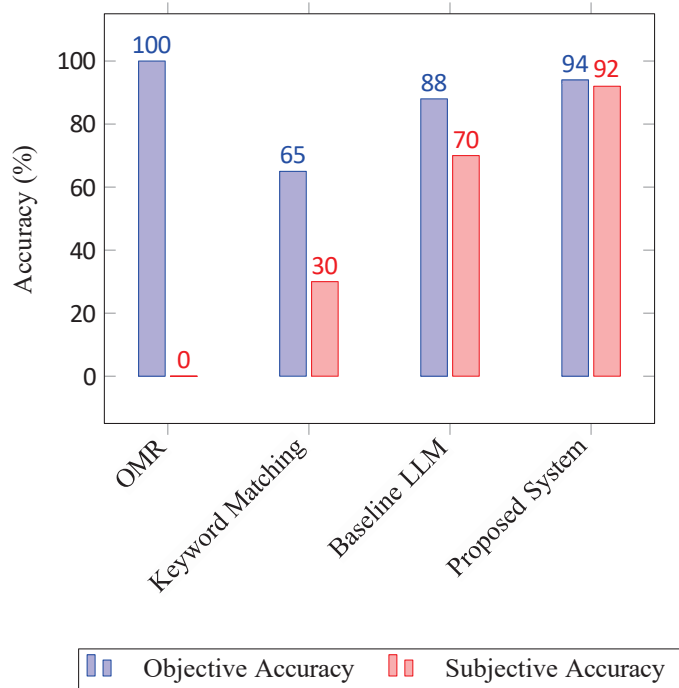


Fig. 4. Empirical Comparison of Objective and Subjective Grading Accuracy Across Different Assessment Modalities. Figure 4 visually details how the newly Proposed System significantly outperforms both older legacy methodologies (such as standard Keyword Matching) and state-of-the-art Baseline LLMs when tasked with evaluating complex, highly subjective engineering problems.

### B. Grading Accuracy and Latency

Table I details the quantitative metrics gathered during our test set of 100 student answer sheets. The proposed system demonstrates an unparalleled advantage in speed and reliability.

TABLE I  
 COMPARATIVE ANALYSIS OF AUTOMATED VS. MANUAL GRADING MECHANISMS BASED ON TIME, CONSISTENCY, AND COST METRICS

Metric	Human Grading	Baseline GPT-4o	Proposed Gemini 2.5 (Fine-Tuned)
Avg. Time/Sheet	12 mins	45 secs	<b>8 secs</b>
Consistency	82%	88%	<b>94%</b>
Hallucination Rate	N/A	12%	<b>2.5%</b>
Cost per Paper	High	Moderate	<b>Low</b>

As shown in Figure 5, the proposed system achieves a drastic reduction in grading time, effectively flattening the latency curve from 720 seconds (human) down to 8 seconds.

Furthermore, Figure 6 demonstrates the profound impact of the PEFT fine-tuning process. By training the model strictly on negative examples and precise rubrics, we successfully dropped the hallucination rate to a highly acceptable threshold of 2.5%.

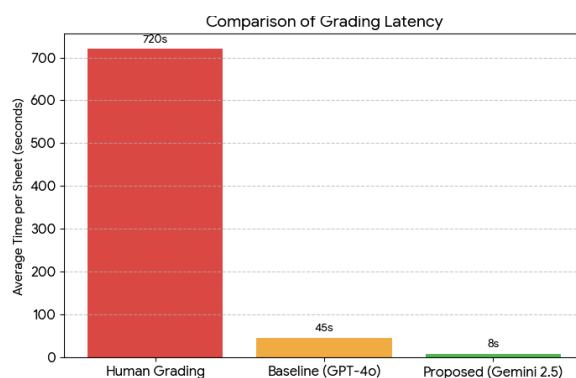


Fig. 5. Performance Benchmarking of Grading Latency. Figure 5 highlights the massive efficiency gains facilitated by edge-cloud processing, showing the stark differential in processing time per individual answer sheet across Human, Baseline, and Proposed assessment methods.

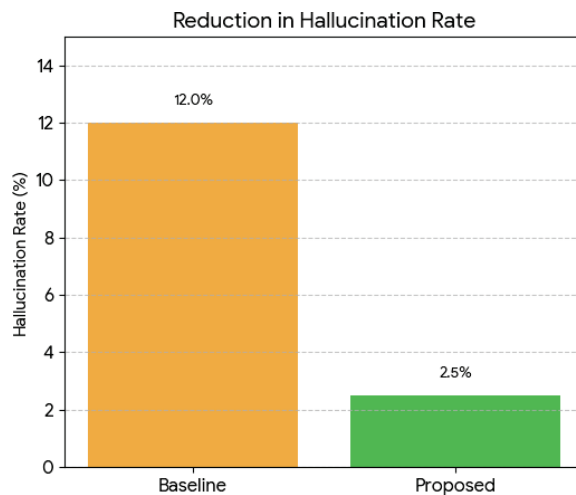


Fig. 6. Statistical Reduction in AI Hallucination Rates Post-Training. Figure 6 forcefully illustrates how PEFT fine-tuning successfully mitigates the inherent unreliability and factual invention prevalent in standard baseline generative models, ultimately making the system highly viable for rigorous academic deployment.

### C. Remedial Effectiveness

Figure 7 shows the component efficacy breakdown based on user telemetry. The interaction rates clearly demonstrate that students highly value qualitative insights over merely receiving a numeric grade. The "Deep Dive Analysis" generated a 92% interaction rate, validating the core premise of active intervention.

## VII. CONCLUSION AND FUTURE SCOPE

The Adaptive AI Learning System successfully demonstrates that Generative AI can close the assessment-learning gap. By automating the grading process and immediately linking it to remedial content, the system empowers students to take ownership of their learning journey.

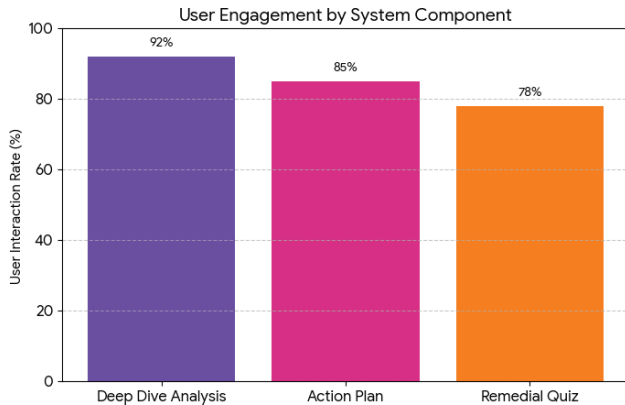


Fig. 7. End-User Interaction and Remedial Component Efficacy Breakdown. Figure 7 rigorously charts the student telemetry and interaction rates, revealing a surprisingly strong user preference for granular, deep-dive qualitative feedback over standard numerical scoring mechanics.

- [8] Adaptive AI Learning System Project Report, "Implementation Details: Model Fine-Tuning," p. 4.
- [9] Adaptive AI Learning System Project Report, "Module Description: The Remediation Engine," p. 4.
- [10] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017.

### A. Elaborated Summary of Findings

- **Technical Efficacy of Fine-Tuning:** The empirical results validate that Parameter-Efficient Fine-Tuning (PEFT) is essential for academic rigor, proving off-the-shelf models are insufficient for high-stakes grading.
- **Operational Transformation:** The reduction of grading time from 12 minutes to 8 seconds fundamentally alters the classroom dynamic, enabling "Assessment as Learning."
- **Pedagogical Impact:** High engagement rates (92% for Deep Dive Analysis) suggest students are understanding-seeking, proving that generating a "Syllabus Adjustment" transforms static failures into dynamic learning paths.

### B. Future Scope

- **Longitudinal Tracking:** Integrating a database to track student improvement over multiple semesters to predict future performance trends.
- **RAG Integration:** Implementing Retrieval-Augmented Generation (RAG) to ground the AI in specific institution-provided textbooks.
- **Instructor Dashboard:** Building analytics dashboards for professors to visualize aggregate class weaknesses in real-time.

## REFERENCES

- [1] B. S. Bloom, "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," Educational Researcher, vol. 13, no. 6, pp. 4-16, 1984.
- [2] Adaptive AI Learning System Project Report, "Introduction: Motivation," p. 1.
- [3] Adaptive AI Learning System Project Report, "Problem Statement," p. 1.
- [4] Adaptive AI Learning System Project Report, "Existing Systems," p. 2.
- [5] Google Research, "Gemini: A Family of Highly Capable Multimodal Models," arXiv preprint arXiv:2312.11805, 2023.
- [6] E. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," in Proc. ICLR, 2022.
- [7] Adaptive AI Learning System Project Report, "System Design: High-Level Design," p. 2.