Preliminary Report CM3070 Final Project:

An Offline AI Education Assistant

Template 3.1: Orchestrating AI models to achieve a goal

1. Introduction

1.1 Overview

The goal of my project to develop a multimodal AI powered tutor to help students solve homework problems and help in their academic growth. The project falls under the template, Orchestrating AI models to achieve a goal. According to the requirements of the template we do not have to train any model ourselves from scratch and have to use already available pre-trained models combined to meet a unified, clear purpose. The pre-trained models should also work on different modes of input such as images, audio and text. The central concept of this project template requires the integration and orchestration of specialized AI models, each with distinct computational requirements and data processing methods. Moreover, the template explicitly emphasizes the importance of clear, purposeful integration rather than isolated usage of these models.

This assistant will interpret homework questions submitted as text, image, or audio. Then provide step by step, coherent explanations through text or speech outputs. By integrating Optical Character Recognition (OCR), Speech to text (STT), Translation models, Text to Speech (TTS), and a large language model (LLM), this assistant is intended to facilitate learning in diverse and low resource educational environments. These environments can include public schools in developing regions or homes without stable internet where access to modern AI tools is currently limited.

1.2 Motivation

In recent years, AI has significantly transformed education and enhanced learning. According to a controlled study [1] of college students it was found that students learned twice as fast as compared to active learning in class, while also being more engaged and motivated. A meta-analysis of 51 studies on ChatGPT in education found that it has a substantial positive effect in a student's learning performance, higher-order thinking and engagement while also reducing cognitive load [2]. AI can not only help in students carrying out their homework but also enhance their overall educational experience and outcomes.

Advanced models such as OpenAI's GPT-4 and DeepSeek's large language models have demonstrated impressive capabilities in educational assistance, making complex subjects accessible and engaging for students worldwide. However, these powerful AI tools often require persistent internet connections, substantial computational resources, and can incur high ongoing subscription costs, making them unsuitable for public schools, rural communities, and lower-income households.

Cloud based AI solutions also pose significant privacy concerns, as student data and educational content are typically processed on remote servers. This raises questions of data sovereignty, security, and compliance with privacy regulations such as GDPR and FERPA. As a result, there's a clear and pressing need for a privacy centric, accessible, and fully offline AI educational assistant that addresses these concerns.

Furthermore, multilingual support remains limited in mainstream AI education tools, particularly for languages like Urdu, which serves millions of users and ranks tenth worldwide in number of speakers [3]. Students in regions such as Pakistan, India, and Bangladesh often encounter language barriers, limiting their ability to fully leverage digital educational resources. Addressing these multilingual needs through a locally deployable system can significantly reduce educational disparities and improve learning outcomes for these underserved student populations.

In Pakistan, a significant number of schools lack internet access. Only 21% of urban government schools and 40% of private schools have internet access. In rural areas the situation is even worse with only 14% of rural government schools and private schools being equipped with internet facilities [4]. Teacher shortages exacerbate the problem even further. 30% of government run schools are single-teacher schools which means there is only one teacher for classes 1 through 5 [5]. Gender disparities persist across all levels of education in Pakistan, with boys outnumbering girls at every stage [6]. One underlying factor is the cultural norm that restricts girls' mobility and discourages them from interacting with boys, thereby limiting their access to schooling.

Parents in Pakistan often keep daughters at home when the nearest school is co-ed, far away, lacks female staff, or has poor sanitation, fearing harassment and social disapproval. Limited resources also lead families to prioritize sons, viewed as future earners, while girls are expected to handle domestic roles or marry early. These norms and barriers deter girls from enrolling and staying in school, widening the gender gap at higher grades.

1.3 Aims and Objectives

This project explores how comparable functionality can be achieved completely offline using lightweight, open-source models that run on local machines. The goal is not to match the full capability of systems like GPT-4o, but to deliver a cost-free, private, and a highly accessible alternative that can be installed and run on low-end laptops, shared school computers and devices like the Raspberry Pi 5.

The assistant is compliant with the data protection laws, GDPR and FEPRA as all processing happens locally. Parents or eligible students keep full control of their data. They can view, correct or delete any session data with a single click, fulfilling the right of access/erasure in both regulations. In practice this privacy native architecture lowers adoption barriers for public schools and removes the need for complex data processing agreements.

The primary goal of this project is to design and evaluate a system that allows students to submit images (of handwritten or printed homework questions), speech, or typed queries. The assistant uses OCR/STT/Translation models to convert inputs into a structured format then sends the converted input to a language model for explanation. It returns a step-by-step answer optionally in spoken form via TTS and works completely offline, without any server or internet requirement.

Deploying a fully offline AI tutor on low-end hardware like the Raspberry Pi can significantly mitigate the challenges posed by limited teacher availability and lack of internet access in schools and homes. Girls unable to attend school due to cultural and social norms in Pakistan and South Asia can also benefit substantially from this project. The AI tutor provides an accessible form of homeschooling, potentially encouraging continued education and helping reduce gender-based educational disparities.

2. Literature review

2.1 Effectiveness of AI in Education

AI has significantly reshaped educational methodologies and learning environments over the past several decades. Historically, the educational application of AI began with the development of Intelligent Tutoring Systems (ITS). A software designed to replicate the personalized instruction typically provided by human tutors. Early ITS primarily employed rule-based logic to diagnose student misunderstandings and offer tailored guidance. A review an educational psychologist [7] compared the effectiveness of human tutors, ITS, and conventional instruction. His comprehensive analysis showed that ITS often-outperformed traditional classroom teaching methods and were nearly as effective as one-to-one human tutoring, highlighting the potential for personalized AI-driven education.

AI began shifting from traditional rule-based ITS toward machine learning (ML), and specifically deep learning-based approaches. Neural network based tutoring systems started leveraging large scale data analysis to predict learner behaviors, identify misconceptions, and personalize content dynamically [8]. This shift was accelerated by breakthroughs in Natural Language Processing (NLP), computer vision, and speech recognition, facilitating richer interactions between students and AI agents. Rather than relying solely on fixed rules, neural models could now analyze patterns in large datasets, adapt to individual learning styles, and continuously refine their instructional strategies based on new data.

Recent empirical evidence strongly supports AI-assisted learning's superiority over traditional classroom methods [1]. Findings from several meta-analyses of AI educational interventions consistently demonstrate enhanced cognitive outcomes, improved student engagement, and reduced cognitive load when students learn with AI-powered tutors compared to conventional methods [9].

2.2 Existing Similar Projects

Chat GPT:

ChatGPT-4o (OpenAI, 2024) is a multilingual, multimodal AI. It offers interactions via text, voice, and image inputs and outputs. Studies have shown that AI tutors, like ChatGPT, can significantly enhance student engagement, motivation, and learning outcomes, especially in higher-order thinking and problem-solving skills [10]. ChatGPT is free but paid subscription plans have higher usage limits.

However, ChatGPT is limited by its reliance on cloud infrastructure, ongoing subscription costs, and significant privacy concerns related to student data processing on external servers. It also lacks consistent multilingual support especially for less commonly supported languages such as Urdu. This limits its practical use in educational contexts particularly in rural and underserved regions.

KhanMigo

Khanmigo is developed by Khan Academy and is based OpenAI's GPT-4. It is an education support AI and provides tailored instructional explanations and structured step-by-step guidance. KhanMigo enhances personalized student engagement and learning outcomes. Its effectiveness has been validated through its integration in classroom environments. Khanmigo shares similar limitations with ChatGPT-4o, including dependence on cloud services, stable internet access, privacy concerns, and limited multilingual capability.

Photomath (BYJU'S)

Photomath is a specialized educational assistant designed primarily for mathematics. It excels in processing handwritten mathematical problems through advanced OCR techniques. It offers detailed, step-by-step animated explanations. Its intuitive interface and effective handwriting recognition have made it popular among students needing math support. Photomath is available as a smartphone app and owned by Google. Photomath is limited due to its cloud dependent nature and lack of support for other subjects or multimodal input/output formats.

Intelligent Tutoring Systems (ITS)

Intelligent Tutoring Systems represent another category of educational AI technology. They focus heavily on instructional and adaptive learning paths. Prominent examples include Carnegie Learning's MATHia (Cognitive Tutor) and AutoTutor. MATHia has demonstrated strong empirical evidence of improving student learning in mathematics through structured, adaptive problem-solving exercises [11]. AutoTutor, developed to facilitate dialogue-based tutoring. It engages students through natural-language conversation, effectively producing learning gains [12].

While these ITS have proven educational efficacy, they often require extensive infrastructure, ongoing maintenance, specialized installations, and lack multimodal, multilingual, and offline functionalities. For example, MATHia is web based or desktop dependent and requires significant licensing fees, whereas AutoTutor usually relies on online servers, limiting its deployment in resource-constrained regions. Other ITS platforms also share these limitations, being predominantly math focused and cloud dependent.

Offline and Edge Computing Educational Initiatives

There have been some relevant attempts to deploy AI and digital education solutions offline. Kolibri, developed by Learning Equality, is a notable example. It provides open-source offline content delivery to low resourced schools. Studies reveal that offline, localized educational technologies such as Kolibri significantly improve learning accessibility and engagement in disconnected communities. However, Kolibri lacks sophisticated intelligent tutoring capabilities, limiting itself to content delivery rather than interactive personalized guidance [13]. However, they are still limited in functionality and capabilities, such as kKolibri is only a content delivery system.

2.3 Current Project

The reviewed literature clearly indicates that while AI educational technologies have made remarkable improvements limitations still remain. Major concerns being internet dependency, inadequate multilingual support, and data privacy. Current cloud-based tools and ITS solutions tend to prioritize richer user experiences and adaptive educational personalization but fail to support students in offline, multilingual, or economically disadvantaged settings. This project is designed to address these limitations through an orchestrated, edge-friendly system.

All models (OCR, STT, LLM, TTS) run locally eliminating third party data transfer, thus aligning with GDPR and FERPA privacy regulations. This contrasts sharply with cloud systems such as ChatGPT, Photomath, and Khanmigo, whose privacy models rely on contractual safeguards rather than technical isolation. This project can be used in schools while avoiding complex data processing agreements which extends its educational impact and usability.

By combining the models, the assistant handles handwritten/typed images, speech input, and Urdu–English text, all on CPU. The ITS or commercial assistant reviewed do not provide this extensive modality and bilingual support without internet. This project promises practical educational benefits in regions previously neglected by educational technology.

Existing offline initiatives (e.g., Kolibri) deliver static content; they lack dynamic, step-wise reasoning. Using a 1.5 B-parameter LLM distilled from DeepSeek-R1 demonstrates that chain-of-thought explanations can be generated under 8 GB RAM, a technical contribution to edge-AI literature where most reasoning benchmarks assume >7 B models and GPU resources. Recent research on edge computing and lightweight AI models deployed on devices like Raspberry Pi illustrates the technical feasibility of deploying advanced AI models offline [14]. These projects generally highlight low computational power as a primary challenge constraining model complexity and capability. Despite these limitations they strongly support the feasibility of the current project's hardware strategy and indicate clear opportunities to leverage quantization techniques to balance model capabilities and performance on low end hardware.

All chosen components and models are open source. Students incur zero recurring fees, directly countering subscription barriers stated for AI assistants such as ChatGPT. The project therefore contributes a replicable, financially sustainable system for AI adoption in the education sector.

Many AI assistants are not designed to be a tutor. Through prompt engineering techniques the LLM used would be tuned to act a tutor and provide scholastic learning, such as asking student related questions before giving a full answer. The AI assistant can also be prompted to ask student questions to ensure they have fully learned what has been taught.

Feature	Proposed AI assistant	ChatGPT	KhanMigo	Photomath	MATHia (ITS)
Input Modalities	Text, Image (printed + handwritten), Voice	Text, Image, Voice	Text, Image, Voice	Image	Interactive problem sets
Output Modalities	Text, Speech (offline TTS)	Text, Image, Voice	Text, Speech	Animated math steps	Step-by-step hints
Offline Capability	Yes	No	No	No	License requires periodic sync
Multilingual	English and Urdu, expandable	Partial, inconsistent	Limited	Limited	English Only
Data Privacy	All processing local (GDPR / FERPA compliant)	Cloud data storage	Cloud data storage	Cloud data storage	Cloud / institutional servers
Cost Model	Free, open source.	Subscription / API fees	School subscription	Freemium / BYJU'S+	Per-seat license
Hardware Requirement	8GB RAM and CPU only	Datacenter GPUs	Cloud	Smartphone and cloud	PC and server backend

2.4 Limitation and Mitigations

Despite its advantages the offline AI tutor faces several constraints that can be however, mitigated. First, CPU only inference of a 1.5 billion parameter model on devices such as a Raspberry Pi yields slow generation (roughly 4 tokens/s. to overcome this frequently asked questions (FAQ) and concept explanations can be pre-generated and stored locally for instant retrieval. Where a desktop is available a larger model can be used. Second, the smaller model's reasoning depth is capped; therefore, the project is scoped initially to primary and middle school. The project is designed so a 6.7 B or 7 B GGUF file of the model can be swapped in without code changes. As a fallback tiny language models can also be opted to improve performance gains.

Handwriting OCR is error prone in low-contrast images. Images can be run through both handwritten and printed models; higher confidence output will be chosen. OpenCV pre-processing and a user-editable text preview can be provided to prevent errors. Language coverage is presently limited to English-Urdu but this can be easily mitigated by adding additional language translation models.

Given the project's tight timeline, I will prioritize and implement only the most critical mitigation measures rather than an exhaustive set of enhancements. However, I will try to attempt them all.

2.5 Potential Impact

The current project offers a practical demonstration that high-quality, multimodal, bilingual tutoring can be delivered entirely offline on commodity hardware. It bridges a documented equity gap in rural and low-income regions [13], provides a replicable architecture for further research, and uses open-source tools that policymakers and NGOs can deploy at a larger scale. It has the potential to advance both the academic discourse on edge AI and the real-world mission of inclusive education.

3. Project Design

3.1 Overview

The project falls under the template: Orchestrating AI models to achieve a goal. This project integrates different pretrained models and runs them offline on a low cost and resource hardware to create an AI tutor. Students can photograph handwritten/printed questions, speak their queries, or type them. The AI assistant returns a step-by-step explanation in English or Urdu text which can be optionally read aloud. All inference is performed locally by orchestrating five pre trained models in the domain of:

1. Transformer based OCR
2. Speech to text model
3. Large Language Model
4. Text to speech model
5. English – Urdu translation model

The project's architecture is deliberately modular, allowing any model to be swapped or upgraded as needed.

3.2 Domain and User

This project combines educational technology, natural language processing, and edge AI. Its specific niche is offline, multimodal homework assistance for primary and middle school students (grades 1–8) in environments where reliable internet, subject-specialist teachers, and commercial AI subscriptions are unavailable or expensive. The project is suitable for any student up to middle-school level, but its primary target users are learners in low-resource environments.

Primary Users – Rural and Low-Income Students.

These pupils attend government or under resourced private schools where a single teacher may handle all subjects for multiple grades. Because reliable internet is rare cloud-based AI tools are unusable. The offline tutor offers them instant, step-by-step help in English or Urdu, deployed on a shared classroom PC or a Raspberry Pi kiosk at the school. For girls who are kept at home due to cultural and social constraints, the same software can run on a family laptop and provide a form of private, self-paced homeschooling. This gives them a viable pathway to continue education.

Secondary Users – Overburdened Teachers, Parents, and Self-Directed Students.

Single-teacher classrooms struggle to provide personalized attention, and time is spent on repetitive explanations of fundamental concepts. By delegating routine Q&A to the AI tutor, teachers gain time to focus on more complex tasks like project work, formative assessment, and classroom management. Parents with limited subject knowledge can also utilize the assistant at home to scan their child's worksheet and receive a clear step-by-step solution. This allows them to supervise homework without becoming content experts themselves. All processing is done locally, so teachers and parents avoid data privacy paperwork and ongoing subscription fees. A growing cohort of tech-enabled students already have access to computers can run the assistant independently to review material, practice problem solving, and fill knowledge gaps outside class hours.

Tertiary Stakeholders – Education Officials, NGOs, and Researchers.

District education officers and non-profits organizations need scalable, low-cost solutions that improve learning outcomes without new infrastructure. Because the tutor is open-source, runs on relatively inexpensive hardware, and complies with GDPR/FERPA by design, it can be rolled out across schools with minimal hurdles. For academic researchers, the project provides a replicable, edge-AI architecture and a field laboratory for studying multimodal tutoring in low-resource contexts, filling a gap in current literature focused on cloud-centric solutions.

In essence, the domain is resource constrained primary education, and the users are students, teachers, and parents seeking an affordable, privacy-safe, language-appropriate AI tutor that does not require internet access. By combining privacy first design, low hardware requirements, and bilingual support, the AI tutor aims to provide support to overstretched single-teacher classrooms and enables home based learning. The success of this project will demonstrate that high quality AI tutoring is achievable on low spec hardware, laying a foundation for scalable and open-source educational interventions in resource constrained environments.

3.3 Design Justification

The limitations found in the users and domain are reflected in the fundamental design approach, which is to orchestrate pretrained models in a modular, offline manner. Any cloud-based design would leave out the vast majority of the intended beneficiaries. Operating OCR, STT, LLM, translation, and TTS locally on a Raspberry Pi or repurposed laptop ensures continuous availability in homes and classrooms without connectivity or with limited financial resources.

User variety requires a multimodal input pipeline (camera, microphone, and keyboard). Younger kids or parents with low literacy can ask inquiries. Students with handwritten/printed notes can photograph work and confident typists can simply type text. This three-input design is consistent with cognitive load literature, which shows that learning improves when students interact through their preferred modality [15], while also accommodating the hardware variability common in under resourced schools (for example, some classrooms have webcams but no mics, others the opposite).

Bilingual support (English - Urdu) is not an optional feature, but rather a direct response to field data indicating that comprehension suffers significantly when explanations are delivered solely in English [16] rather than in the mother tongue. Integrating a translation model before and after the LLM enables the same reasoning engine to serve both language communities without the need for two separate models, reducing RAM usage and simplifying updates.

Running the AI assistant locally follows a Privacy by Design (PbD) approach, which ensures compliance with GDPR and FEPRA regulations. This eliminates the need for schools to negotiate data processing agreements and reassure parents who are concerned that their children's voices or worksheets will be uploaded to unknown servers. This architectural design also increases the adoption rate of such a system.

Finally, a loosely linked, swappable component layout helps to future proof the system. If testing reveals that the 1.5 B DeepSeek model is insufficient for tasks, a quantized 7 B DeepSeek GGUF can be substituted without affecting the OCR, translation, or UI layers. Similarly, regional NGOs could replace Urdu translation with Pashto or Sindhi by simply substituting the translation model. This extensibility ensures that the project remains relevant as curricular demands, hardware budgets, and open-source models evolve.

3.3 Project Architecture

The project is organized as a layered, modular stack that runs entirely on a Raspberry Pi 5 or any low spec laptop with Python 3.10. Tkinter user interface will be a single window which presents a chat like answer pane, a multiline text box for typed questions, and two buttons, Upload Image and Record Audio. Because Tkinter ships with every standard Python build, no additional GUI dependencies are required. When a pupil submits an image, speaks a query, or types directly, the UI drops a task into a thread-safe queue and immediately returns to an idle state. The processing occurs off the main thread so the window never freezes.

A small controller module retrieves that task, identifies the modality, and routes it to the appropriate pre-processing pipeline. Images pass through two transformer-based OCR models, TrOCR-Printed and TrOCR-Handwritten, and the controller keeps whichever string reports the higher confidence score. Audio clips are transcribed by Whisper-Small, chosen for its robustness to classroom noise; both OCR and speech recognizer run entirely on CPU. If the language detected is Urdu it is passed through OPUS-MT to translate to English, ensuring that the LLM always receives an English input.

Reasoning is handled by a 1.5 billion parameter DeepSeek R1 Distill Qwen model stored. The controller wraps the pupil's question in a tutor style system prompt ("You are a patient homework tutor. Explain step-by-step …") and streams the generated tokens back to the UI thread as they arrive, giving students visual feedback even if CPU generation is slow. The raw answer is optionally translated back to Urdu, and if the speaker toggle is active the raw answer is routed to a text to speech model depending on the language.

Because every functional block lives in its own Python module (ocr.py, stt.py, translate.py, llm_engine.py, etc.), the design is swappable. A school with a desktop PC can change to a 7 B or 14 B DeepSeek. the controller will reload the larger model at the next query with no other code changes. Likewise, adding Pashto support is as simple as adding a new translation model and an option to switch to the language.

3.4 Work Plan

The table below shows the project's proposed timeline. The timeline is flexible, with work moving on to the next feature once the current one is completed. Towards the end of the project (weeks 18-19), additional time is set aside for each central feature to accommodate any changes that may arise after user testing. The goal is to complete the project by week 20. The remaining three weeks before submission will serve as a buffer against unexpected delays or to enhance the project once it reaches MVP status.

Week	Milestone	Deliverable
W2	Template finalized	Finalize and research on project template
W4	Literature Review	Literature Review and research existing solutions
W6	Project Design Complete	Architecture, technology stack, work-plan
W8	Feature Prototype	Implement basic feature prototype
W10	Preliminary Report Submission	Prototype and draft submission
W12	UI and OCR pipeline	Image → OCR → LLM → English text displayed in Tkinter
W15	Multimodal Prototype Ready	Speech input, Urdu translation, English & Urdu TTS integrated
W17	Edge-Device Performance Target Met	Ensure workable performance on a Raspberry Pi 5
W18	Testing	SUS and user testing
W20	Project finalize	Final touches and bug fixes
W21	Final Report	Complete project submission
W22	Additional Features	Additional features for increased usability

3.5 Contingency Plan

Although the schedule leaves buffer time in every phase, several technical and logistical risks could still jeopardize the delivery of a fully-functional prototype. The table below outlines the most likely failure modes and the pre-approved fallbacks that can be deployed without derailing the project timeline.

If DeepSeek-1.5 B proves too slow then swap to TinyLlama-1.1 B for Pi builds but keep DeepSeek for laptop edition. If Whisper STT lags in real time, then fall back to WebRTC client. If Urdu TTS training data quality is poor, then default to English voice and display Urdu text subtitle.

With this schedule the most critical risk, multimodal pipeline integration, occurs early, while ample buffer time remains for optimization, testing, and documentation ahead of the September submission.

3.6 Evaluation Plan

To determine whether the prototype truly fulfills its offline tutoring goal, the project will be evaluated on three categories: technical performance, educational impact, and usability. Each category has its own set of instruments, success criteria, and corrective actions, ensuring that the evaluation is both comprehensive and in line with the project's objectives.

Technical Performance Evaluation

Before any classroom exposure, the system must demonstrate that it can run acceptably on target hardware. An automated Python harness will feed a 50-item mixed-modality test set (25 printed images, 10 handwritten images, 10 audio clips, 5 typed questions) through the pipeline while psutil logs latency and peak RAM. The prototype passes this test only if it:

- Returns an image-to-speech answer in < 8 s on a Raspberry Pi 5 (8 GB)
- Keeps peak RAM below 7.5 GB
- achieves > 90 % OCR character accuracy on printed text
- Maintains speech recognition word error rate < 20 % in moderate classroom noise.

Any shortfall triggers optimization or a model swap (e.g., tighter quantization) before user testing proceeds.

Educational Impact Evaluation

Because I only have five student volunteers (ages 9-13), educational impact will be tested with a very small study that can be run in a single afternoon. First, I will prepare a worksheet of six grade-level questions arranged in two matched sets: three items in Set A and three parallel items of equivalent difficulty in Set B. Each learner will attempt Set A without any assistance for ten minutes, then, after a short break, solve Set B using the AI tutor. For both sets I will record the number of correct answers and the time taken. Immediately after finishing Set B, the student will rate the tutor on two five-point scales: (1) "How helpful was the explanation?" and (2) "Was the language easy to understand?". The tutor will be considered educationally effective if, when averaged across the five participants, it produces at least one more correct answer per learner, matches or improves the median completion time, and receives an average rating of 4 out of 5 or higher on both perception questions. Although small, this paired comparison provides concrete evidence that the offline assistant improves accuracy, maintains or even accelerates problem-solving, and provides explanations that young learners find clear and useful.

Usability Testing

Because the target audience includes young learners and low-tech teachers, ease of use is critical. Directly after the educational testing, the same participants will complete the System Usability Scale (SUS). A mean SUS score of 70 or above (the threshold for "Good") will be considered acceptable. Think-aloud observations captured during the testing will supply qualitative data for interface refinements; any severe pain points will be addressed in the Week 20.

4. Feature Prototype

The first feature prototype already demonstrates the complete offline flow from a learner's input to a step-by-step answer. A user can either type a homework question directly in the console or supply a photo of a printed worksheet. The program then recognizes any printed text in the image, feeds the resulting question into a local language model, and returns a coherent explanation that is suitable for a middle-school pupil—all without any internet connection or external server.

Technically, the prototype replaces the earlier TrOCR experiment with **EasyOCR**, a lightweight convolutional-recurrent network that runs entirely on CPU and accurately extracts English text from photographs. The recognized sentence is passed to a 1.5-billion-parameter **DeepSeek-R1 Distill Qwen** model, which is loaded from safetensors in half-precision and executed on the CPU. A small helper function strips the prompt so that only the answer appears in the console. EasyOCR returns text in about a second, while the language model generates an explanation in roughly twenty-five seconds, well within an acceptable wait time for a homework setting.

Several limitations remain. The current OCR path struggles with messy handwriting, the response latency could be halved by quantising the language model to 4-bit weights, and the system is still English-only. The next development sprint will integrate a handwriting-optimised TrOCR model as a fallback, add MarianMT translation and Piper Urdu text-to-speech modules. Nevertheless, the working prototype already proves the core claim of the project: high-quality, multimodal tutoring can run entirely offline on commodity hardware.

References

1. Gregory Kestin*, Kelly Miller*, Anna Klales et al. AI Tutoring Outperforms Active Learning, 14 May 2024, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-4243877/v1]

2. Wang, J., Fan, W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanit Soc Sci Commun* **12**, 621 (2025). https://doi.org/10.1057/s41599-025-04787-y

3. Zeidan, Adam. "Languages by total number of speakers". *Encyclopedia Britannica*, 7 Aug. 2023, https://www.britannica.com/topic/languages-by-total-number-of-speakers-2228881. Accessed 15 June 2025.

4. ASER Pakistan. 2024. Access to Communication and Technology: Policy Brief. Islamabad: Annual Status of Education Report (ASER) Pakistan. https://aserpakistan.org/document/2024/Access-to-Communication-and-Technology-Policy%20Brief.pdf.

5. Data Stories PK. 2023. "World Teachers' Day Special: Thousands of Teaching Posts Vacant in Schools in Pakistan." Data Stories PK. https://www.datastories.pk/world-teachers-day-special-thousands-of-teaching-posts-vacant-in-schools-in-pakistan/

6. UNICEF Pakistan. 2025. "Education." *UNICEF Pakistan.* https://www.unicef.org/pakistan/education

7. Vanlehn, Kurt. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist. 46. 197-221. 10.1080/00461520.2011.611369.

8. Piech, Chris, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J. Guibas, and Jascha Sohl-Dickstein. "Deep Knowledge Tracing." *arXiv* preprint arXiv:1506.05908 (2015).

9. Yang, Shanshan & Evans, Chris. (2019). Opportunities and Challenges in Using AI Chatbots in Higher Education. 79-83. 10.1145/3371647.3371659.

10. Montenegro-Rueda, Marta & Fernández Cerero, José & Fernández Batanero, José & Meneses, Eloy. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. Computers. 12. 153. 10.3390/computers12080153.

11. Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. Educational Evaluation and Policy Analysis, 36(2), 127–144. https://dx.doi.org/10.3102/0162373713507480

12. Graesser, Arthur & Lu, Shulan & Mitchell, Heather & Ventura, Mathew & Olney, Andrew & Louwerse, Max. (2004). AutoTutor: a Tutor with Dialogue in Natural Language. Behavior Research Methods. 36. 180-192. 10.3758/BF03195563.

13. Shariq Hussain, Zhaoshun Wang, and Sabit Rahim. E-learning Services for Rural Communities. International Journal of Computer Applications, vol. 68, no. 5, pp. 15–20, April 2013. DOI: 10.5120/11574-6888.

14. Merenda, Massimo & Porcaro, Carlo & Iero, Demetrio. (2020). Edge Machine Learning for AI-Enabled IoT Devices: A Review. Sensors. 20. 2533. 10.3390/s20092533.

15. Lehmann, Janina, and Tina Seufert. 2020. “The Interaction Between Text Modality and the Learner’s Modality Preference Influences Comprehension and Cognitive Load.” Frontiers in Psychology 10. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02820 (doi:10.3389/fpsyg.2019.02820).

16. Vacalares, Sophomore. (2023). English-only Versus Mother Tongue: An Analysis on Students’ Fluency and Self-confidence. International Journal of Research and Innovation in Social Science. 07. 1148-1159. 10.47772/IJRISS.2023.7206.