

Component	Description
<p>Background</p>	<p>About the GSMA</p> <p>The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with more than 300 companies in the broader mobile ecosystem. Our mission is to further the strategic, commercial, technological, developing world and advocacy interests of our members. The GSMA produces industry-leading events such as Mobile World Congress, Mobile World Congress Shanghai, Mobile World Congress Americas and the Mobile 360 Series conferences.</p> <p>Every day we work to harness the power of mobile to achieve our industry purpose – Connecting Everyone and Everything to a Better Future – through four main activities:</p> <ul style="list-style-type: none"> • We provide strategic direction and thought leadership to our members • We drive collaborative industry programs aimed at maintaining operator relevance in the digital world • We advocate for the industry to regulators, policymakers and society as a whole • We convene the broad mobile communications ecosystem at our industry-leading events <p>The GSMA INNOVATION FUND</p> <p>The GSMA Innovation Fund is a dedicated team set up to manage all Innovation Fund activities as part of a strategic partnership with the Foreign, Commonwealth & Development Office (FCDO) and The Swedish International Development Cooperation Agency (Sida). The ‘Partnership focuses on Inclusion, Innovation and Scale’, leveraging on 10 years of GSMA Mobile for Development experience running Innovation Funds on a multitude of topics.</p> <p>Orenda</p> <p>There are 22.3 million children in Pakistan out of school, and enrolment and performance both decrease as children move from primary to secondary grades. Orenda bridges the digital divide between private and public schools in Pakistan by providing engaging, high-quality digital education through animated videos and interactive quizzes via its digital platform, Taleemabad.</p>
<p>Project Description</p>	<p>Taleemabad is developing technology to automate the checking of exams via Artificial Intelligence (AI), including homework and in-class assessments, across subjects like Maths, English, and Urdu for primary grades.</p> <p>One of the main challenges is processing Math content, which requires accurate recognition of symbols, equations, and problem-solving steps through Optical Character Recognition (OCR).</p> <p>For Urdu and English, assessment checking is complicated as well. Both subjects present unique difficulties that standard OCR and AI models struggle to handle. There is also a need to ensure that these technologies can work efficiently and accurately across different difficulty levels within each grade.</p> <p>Taleemabad seeks a consultant to support the development of an in-house Generative AI-based solution for automating the checking of Maths, English, and</p>

	Urdu assessments for primary grades. The goal is to solve current challenges in processing these subjects while building long-term expertise and sustainable internal processes.
Key Problem Areas/Challenges	<p>Generic Issues with Exam Checking/Grading</p> <ol style="list-style-type: none"> 1. Strict vs. Lenient Marking: <ul style="list-style-type: none"> ○ In some cases, Generative Pre-trained Transformer (GPT) applies strict grading criteria that differ from teachers' typically more lenient grading style. Conversely, in other cases, GPT may auto-correct students' answers, awarding full marks even when the answers are incorrect. As a result, there is a discrepancy between the scores generated by GPT and those assigned by teachers. ○ Attempted Solution: Providing a marking scheme has been tried to align GPT's grading with the teacher's standards. However, this approach causes the prompt to become too lengthy, which in turn reduces the accuracy of GPT's responses. 2. Skipped Questions: <ul style="list-style-type: none"> ○ While evaluating the exam, GPT sometimes skips certain questions without any identifiable pattern. Although this issue does not occur consistently across all exams, it can happen at any point within an exam. ○ Attempted Solution: Reducing chunk sizes (dividing a student's exam into smaller sections) improved accuracy significantly by ensuring that GPT considers each section more carefully. 3. Incorrect Grading on Non-Text-Based Questions: <ul style="list-style-type: none"> ○ Questions that include elements beyond plain text, such as underlining, circling, or matching lines between columns, often result in inaccurate grading. GPT struggles to interpret these types of questions, leading to marking errors. ○ Attempted Solution: Specific instructions were added to the prompt for these question types. However, accuracy has still fallen short of the required level. <p>Issues with Chunking</p> <ol style="list-style-type: none"> 1. Misinterpretation as a New Exam: <ul style="list-style-type: none"> ○ GPT sometimes mistakenly flags a page as the start of a new exam if it encounters a question like "Introduce yourself" in the middle of an ongoing exam, where students enter their names. This causes GPT to incorrectly interpret the section as the start of a new exam, marking it as "true" instead of "false."

- Identifying the start and end of a test: Currently, we scan all the assessment result sheets in bulk (for a faster upload) which means we need to detect the start and end of the exam.
- Identifying students: We need to read the name accurately and match it with our database. LLMs or OCR tools are unable to read the name effectively in children's handwriting.
- **Current Workaround:** The prompt was modified to avoid flagging as a new exam if student details do not appear at the top of the page.

2. Objective and Subjective Sections Segmentation:

- Many schools format exams so that the objective section is printed and the subjective section is handwritten. Because students often write their names at the start of each section, GPT treats the sections as separate exams. Exam papers may also vary in sequence:
 - Format 1: Objective section followed by subjective section for each student.
 - Format 2: All students' objective sections followed by all subjective sections.
- **Current Workaround:** Instructions were added to the prompt to ignore handwritten sections as separate exams.

Math Exam Checking Challenges

1. Symbol and Equation Recognition: OCR systems often struggle to accurately recognize complex mathematical symbols, equations, and handwritten expressions.
2. Problem-Solving Step Evaluation: Assessing the correctness of problem-solving steps requires understanding logical reasoning and the application of mathematical concepts, which is difficult for AI models.
3. Handwriting Variability: Children's handwriting varies widely, making it challenging for AI models to consistently recognize and interpret their work.
4. Rough work: Confusing the final answer with some of the rough work done on the sheet.
5. Layout management: Understanding the layout of the answers, some equations might be multi-line which might be confusing for the Very Large Memory.
6. Reading and answering questions that involve drawing graphs or diagrams.

	<p>Language Exam Checking Challenges</p> <ol style="list-style-type: none"> 1. Natural Language Understanding: Grading language exams requires understanding nuances in grammar, syntax, vocabulary, and context, which is a complex task for AI models. 2. Following Marking Schemes: Currently adding the marking schemes gives a poor answer. Need to understand why that is. 3. Subjectivity of Assessment: Language assessments often involve subjective judgments, such as evaluating the quality of writing or the effectiveness of arguments. 4. Dialectal Variations: Different dialects and regional variations in language can pose challenges for AI models trained on specific datasets.
<p>Objectives</p>	<ol style="list-style-type: none"> 1. Accuracy and Reliability: The AI system must achieve high levels of accuracy and reliability to ensure fair and consistent grading. 2. Designing a benchmark for the results: How do we effectively measure the accuracy of results? 3. Data Quality: Labelling quality data at scale. We need training data for assessments, questions, answers, and markings. 4. Bias and Fairness: The AI system must be designed to avoid biases and ensure fair assessment of all students. 5. While trying to solve this problem we are hoping to build capability in-house to solve such problems in the future, so there is a core requirement for coaching and mentorship.
<p>Geographic scope</p>	<p>Islamabad, Pakistan. (Remote support is acceptable)</p>
<p>Areas of interest</p>	<p>GenAI, Finetuning.</p>
<p>How it should be done - Approach</p>	<p>To be proposed by the consultant.</p>
<p>Definition of done / final deliverables</p>	<p>Deliverables</p> <p>Project Roadmap</p> <ul style="list-style-type: none"> ● Outcome: Clear, phased development path with milestones, helping Taleemabad allocate resources and track progress effectively. <p>Prototype Development</p> <ul style="list-style-type: none"> ● Outcome: A functional grading prototype that can be tested on sample data, allowing early-stage feedback and iterative improvements. <p>Custom Dataset Guide</p> <ul style="list-style-type: none"> ● Outcome: High-quality, labeled data ready for training, improving AI model accuracy and reliability across question types and subjects.

	<p>Subject-Specific Model Training</p> <ul style="list-style-type: none"> ● Outcome: Models that are fine-tuned to handle complex math symbols, handwriting, and language nuances, enhancing grading accuracy for each subject. <p>Evaluation Framework</p> <ul style="list-style-type: none"> ● Outcome: Benchmarks and metrics to measure AI grading against human standards, ensuring consistency and guiding further model improvements. <p>OCR and Handwriting Recognition Module</p> <ul style="list-style-type: none"> ● Outcome: Accurate identification of student names and exam sections, minimizing errors in student identification and section segmentation. <p>Error Handling Tools</p> <ul style="list-style-type: none"> ● Outcome: Enhanced reliability through tools that identify skipped questions and grading inconsistencies, improving overall grading accuracy. <p>Mentorship and Training Program</p> <ul style="list-style-type: none"> ● Outcome: An in-house team skilled in model customization, prompt engineering, and troubleshooting, empowering Taleemabad to maintain and enhance the system.
<p>The ideal supplier would have</p>	<ul style="list-style-type: none"> ● Expertise in AI Model Development for Education: Extensive experience in AI and machine learning, particularly in educational assessment or similar fields requiring subject-specific customizations, such as Math symbol recognition and language grading. ● Strong Mentorship Skills: A proven track record of mentoring in-house AI teams, especially within non-profit or educational settings. Capable of providing clear guidance on data preparation, model training, prompt engineering, and deployment, to help Taleemabad build sustainable AI capabilities. ● Hands-On Experience with OCR and NLP: Deep familiarity with OCR technologies for complex scripts and symbols (Math, Urdu, English) and natural language processing for subjective assessment grading, ensuring nuanced grading and error handling. ● Proficiency in Data Preparation and Labeling: Demonstrated expertise in designing and managing data-labeling processes, with experience handling handwriting variability and dialectal nuances, to support high-quality training data for primary-grade assessments. ● Practical Experience with AI Deployment: Demonstrated capability in deploying AI models in resource-constrained environments (e.g., offline-

	<p>first, low-power devices) to ensure consistent, reliable performance across varied school settings.</p> <ul style="list-style-type: none"> ● Coding Work (optional): Scope: AI model training, and deployment tasks. ● Time Commitment: 8 hours per week of consultation calls, giving technical support and guidance for the scope above. ● Compensation: \$80/hour 		
Implementation Plan			
Start date	22nd Nov 2024	End date	7 th Feb Jan 2025
Proposed timeline and key steps of exercise*	<p>Tentative Timeline (<i>Timeline is subject to change based on the deployed methodology and complexity of the task</i>)</p> <p>22nd November – GSMA contracting complete</p> <p>2 weeks – Initial Planning and Data Preparation</p> <ul style="list-style-type: none"> ● Conduct initial onboarding sessions and define data collection and labeling requirements. ● Establish evaluation framework and benchmarks for grading accuracy. <p>4 weeks – Prototype Development and Testing</p> <ul style="list-style-type: none"> ● Develop and deploy the first prototype for grading, focusing on OCR and NLP challenges. ● Conduct preliminary testing and gather feedback from Taleemabad's team. <p>2 Weeks – Model Refinement and Subject-Specific Tuning</p> <ul style="list-style-type: none"> ● Refine models based on feedback, with an additional focus on subject-specific grading nuances (Math symbol recognition, language assessments). ● Test and improve the OCR and error-handling modules. <p>1 Week – Final Deployment and Knowledge Transfer</p> <ul style="list-style-type: none"> ● Complete final refinements, integrate models into Taleemabad's system, and conduct performance evaluations. ● Host training sessions and deliver documentation to support long-term in-house management. 		
Risks	<ul style="list-style-type: none"> ● The internal team in Taleemabad not being up to the mark to implement the complexity. ● Lack of accuracy of the models trained for our specific use case. 		

Please submit a proposal to Usman Javed (usman.javed@teleemabad.com) and GSMA Innovation Fund Operations (mjones@gsma.com & uvikram@gsma.com) by the 15th November that includes clear methodology and a step-by-step approach regarding the evaluations of the grantee, including:

- Overall approach and methodology
- A full budget breakdown of all costs
- A full timeline
- Clear evidence of past experience on similar projects

Please also include confirmation of your ability to accept GBP in your bank account (and the country in which your bank account is registered)