

Development of an AI Agent for Automated Educational Question Generation

This report outlines the development of an AI agent that automates the creation of study questions from text. The agent is designed to reason over source documents, then plan and execute the generation of Multiple-Choice (MCQs) and Short-Answer Questions (SAQs). Central to its design is a fine-tuned language model, optimized for this task to ensure high-quality output. The following report shared the evaluation metrics.

Evaluation Methodology:

The AI agent's performance was assessed using a structured human evaluation to quantitatively measure the quality of its generated questions. The methodology was designed to be objective and repeatable.

- **Assessment Scope:** The evaluation covered both Multiple-Choice Questions (MCQs) and Short-Answer Questions (SAQs) generated by the agent.
- **Scoring System:** A 5 point rating scale (1=Poor, 5=Excellent) was used to score each question against a set of predefined criteria.
- **Evaluation Criteria:** The quality of the questions was measured across the following dimensions:
 - Relevance
 - Clarity
 - Answerability from Text
 - Plausibility of Distractors (for MCQs)
 - Specificity (for SAQs)

The final scores in the report represent the average rating for each criterion across all evaluated questions.

Report:

Evaluation Criterion	MCQ Average Score (1-5)	SAQ Average Score (1-5)
Relevance	3	3
Clarity	3.5	2.5
Answerability from Text	4	3.5
Plausibility of Distractors	4	N/A
Specificity	N/A	3.5
Overall Average	3.6	3.1

