PROJECT ABSTRACT

AI Question Paper Generator from Syllabus

A T5 question-generation model trained on SQuAD drafts balanced exam papers from syllabus text, with Bloom's-taxonomy labeling and a marks blueprint.

---

01. Title

AI Question Paper Generator from Syllabus

02. Introduction / Background

Setting a good question paper is slow, skilled work: coverage must be proportional, recall must mix with analysis-level questions, and banks go stale. Faculty do it by hand. This project automates the drafting stage — not the final judgment. From syllabus text, T5 fine-tuned for question generation on SQuAD-derived (context, question) pairs drafts candidate questions per section; a concept extractor ensures coverage; a Bloom's-taxonomy classifier labels each question so the difficulty mix is deliberate; and a blueprint table shows marks across topics and cognitive levels.

03. Problem Statement

Drafting balanced papers manually is slow and coverage is hard to verify, while naive question generators produce repetitive, misleveled questions with no difficulty control. The project addresses this with a controlled pipeline — concept coverage plus Bloom's-level selection against a teacher-set mix — that outputs a formatted draft paper and its blueprint for human review and approval.

04. Objectives

- Fine-tune T5-base on SQuAD-derived (context, question) pairs for question generation.
- Extract key concepts per syllabus topic with TF-IDF + noun-phrase ranking.
- Classify generated questions into Bloom's levels (Remember/Understand/Apply/Analyze) with DistilBERT.
- Select questions to satisfy a teacher-set difficulty mix and total marks with dedup.
- Compute the marks blueprint (topics × Bloom's levels).
- Render a printable draft paper; evaluate generation with BLEU/METEOR plus human ratings.

05. Proposed Methodology

Syllabus text segments into topics; key concepts extract per topic. For each concept, the passage plus concept feed the fine-tuned T5 QG model to draft candidates. DistilBERT labels Bloom's levels. A selector fills the difficulty mix and mark total while maximizing coverage and removing near-duplicates by embedding similarity. The blueprint table computes from the selection; the Flask demo renders the printable paper. Evaluation uses BLEU/METEOR against SQuAD references and human-rated samples.

06. System Architecture / Working

Syllabus → topic segmentation → concept extraction → T5 question drafting → Bloom's classification → mix-constrained selection + dedup → blueprint computation → printable paper + generation summary. The demo runs on a preloaded operating-systems syllabus or pasted text. Output is explicitly a draft requiring teacher review.

07. Hardware / Software Requirements

- Python 3.11, PyTorch, Hugging Face Transformers (T5-base, DistilBERT), scikit-learn
- SQuAD 2.0 (100k+ QA pairs, restructured for question generation)
- Flask + HTML/CSS/JS demo
- Single GPU for fine-tuning; CPU assembly under 10 s per 70-mark paper

08. Expected Outcomes

A generator producing complete draft papers with verifiable coverage and difficulty mix; BLEU/METEOR and Bloom's accuracy reported as achieved figures after training (design targets until then); a demo that turns syllabus text into a formatted paper in seconds.

09. Applications

- Draft paper generation for faculty (with human approval)
- Practice-paper generation for students from their notes
- Question-bank refreshing for coaching institutes
- Teaching text generation + classification in one workflow

10. Future Scope

- Numerical/problem-style question generation (derivations, circuits)
- Difficulty calibration from student performance data
- Multi-subject syllabus handling with subject adapters
- Answer-key and marking-scheme drafting alongside questions