University of Nottingham
UK | CHINA | MALAYSIA

Department of Electrical and Electronic Engineering

A Test and Training Platform Utilizing a Raspberry Pi 5: AI-Retrieval, Tools and a Compact LLM

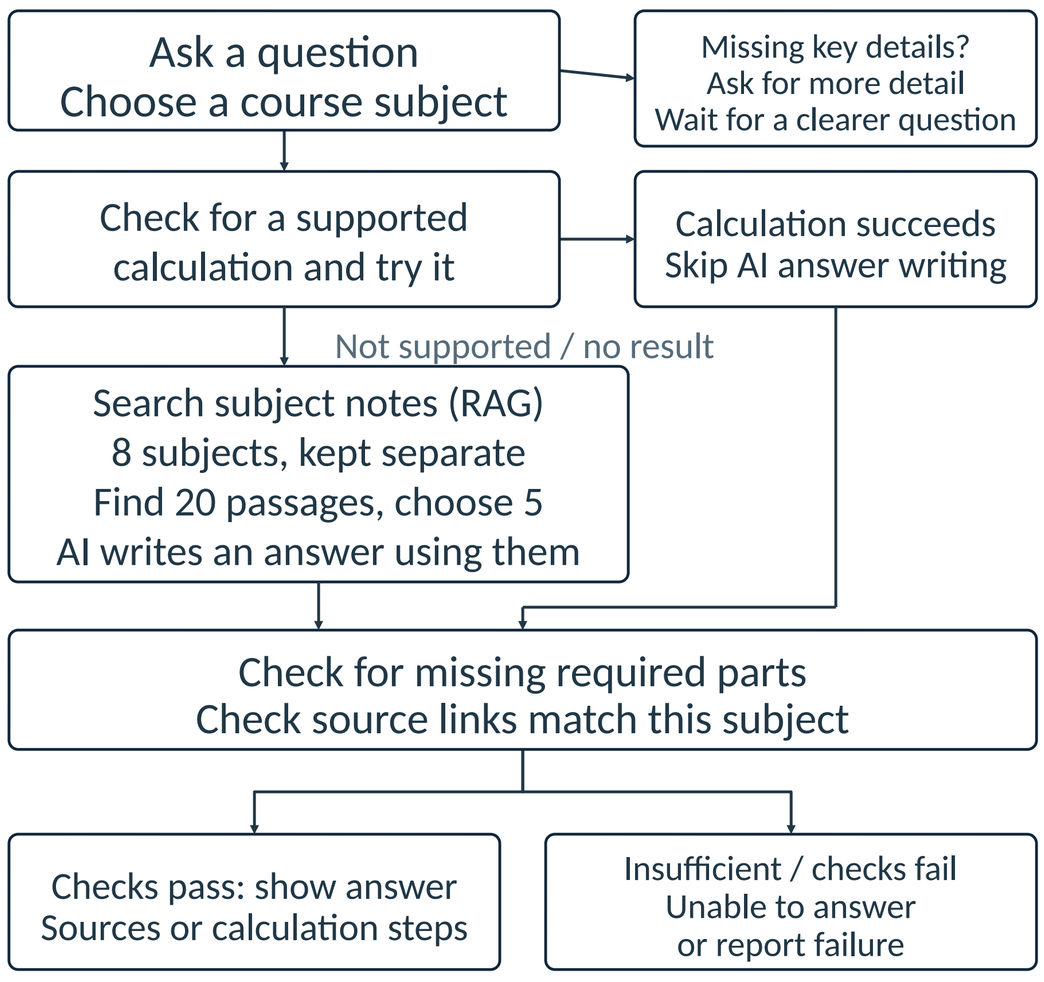
An engineering course assistant for asking questions and practising quizzes.
Model: Qwen3-1.7B.

Target computer



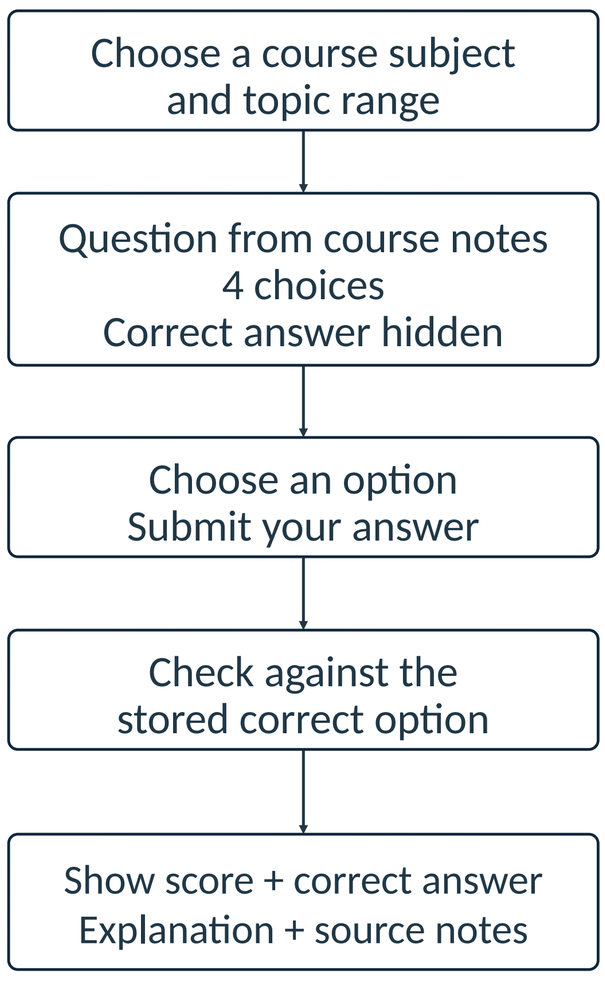
Raspberry Pi 5
16 GB memory / 256 GB storage
Runs Qwen3-1.7B locally
Model compressed to save memory

Ask: questions and answers



These limited checks do not guarantee a correct answer.

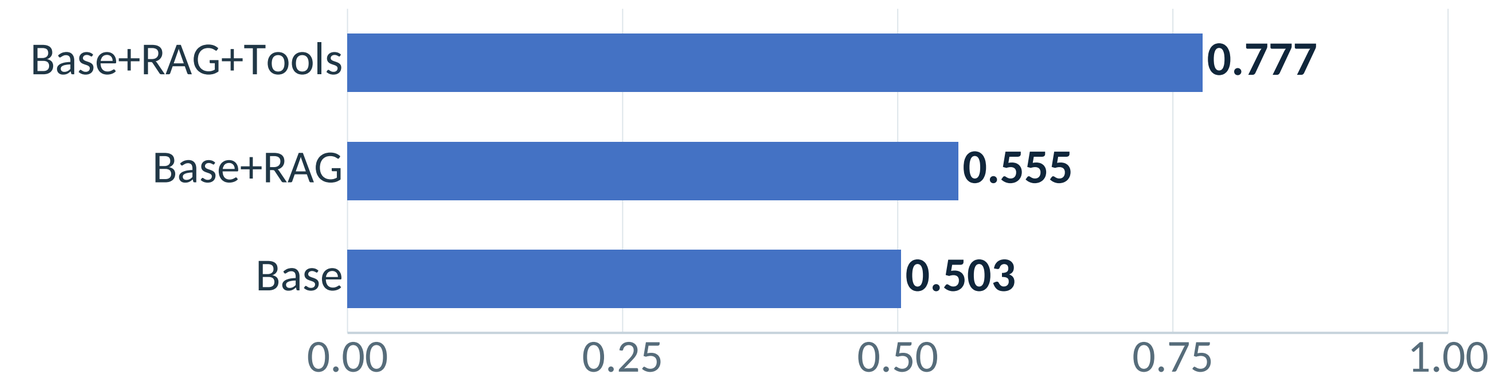
Quiz: practise



RAG (retrieval-augmented generation) supplies course-note evidence before the AI writes an answer.
Offline tests use keyword search (BM25). The Web app searches by meaning (embeddings); this path was not tested.
The tests below assess components separately; Pi tests used AI alone.

Offline answer quality (GPU)

96 questions per setup; 3 of 7 setups shown. Score range 0-1.
Scores assess correctness, completeness, reasoning and relevance.
GPU scores: automatic/AI grading, checked by the first author.



Base = AI alone. Base+RAG = AI with course notes.
Base+RAG+Tools = Calculator if supported, RAG otherwise.

Base+RAG+Tools: average 0.777 out of 1.
Partly correct answers earn points. Fully correct: 54 out of 96.

Separate retrieval test: 24 questions needing course notes.
AI + notes vs AI alone: average score rose by 0.454.
Increase in fully correct answers was not statistically confirmed.

Calculator: 24 preselected supported tasks from the main 96.
Scores rose from 0 to 1 on all 24 versus AI + course notes.

Pi 5: local AI test

AI alone: average answer score (0-1).
Same 96 questions, one answer per device.

0.503 RTX 6000D GPU
0.498 Raspberry Pi 5

Difference in score (Pi-GPU): -0.005.
95% uncertainty range: -0.041 to +0.032.
Equal answer quality is not established.

Median generation speed (tokens/s)

AI alone; 8 questions, repeated requests.
Model already loaded before timing.

317.81 RTX 6000D
8.56 Raspberry Pi 5

Tokens/s = small text units per second.

Pi: 96 AI-alone requests completed.
No retrieval, Calculator or Web quality tests.

Fine-tuning = extra training (GPU)

Two settings: 801 or 2,400 examples.
On 64 questions, neither improved scores over the same setup without extra training.
Data, training or AI instructions may matter.

What remains to be tested

Limited project time kept evaluation internal. We have not tested ease of use or improvements in learning.
Next: independent user tests and quizzes that adapt to topics a learner finds difficult.

Project team: Siheng Ye · Xiaotong Li · Chang Qi · Linfei Yin · Yaokun Zhang · Supervisor: Ian Pratt