

Leaf: Multiple-Choice Question Generation

Kristiyan Vachev¹ Momchil Hardalov¹ Georgi Karadzhov²

Georgi Georgiev³ Ivan Koychev¹ Preslav Nakov⁴

¹Sofia University ²University of Cambridge ³Releva.ai

⁴Qatar Computing Research Institute, HBKU



leafy.me



Paper



SOFIA UNIVERSITY
ST. KLIMENT OHRIDSKI



UNIVERSITY OF
CAMBRIDGE

RELEVA

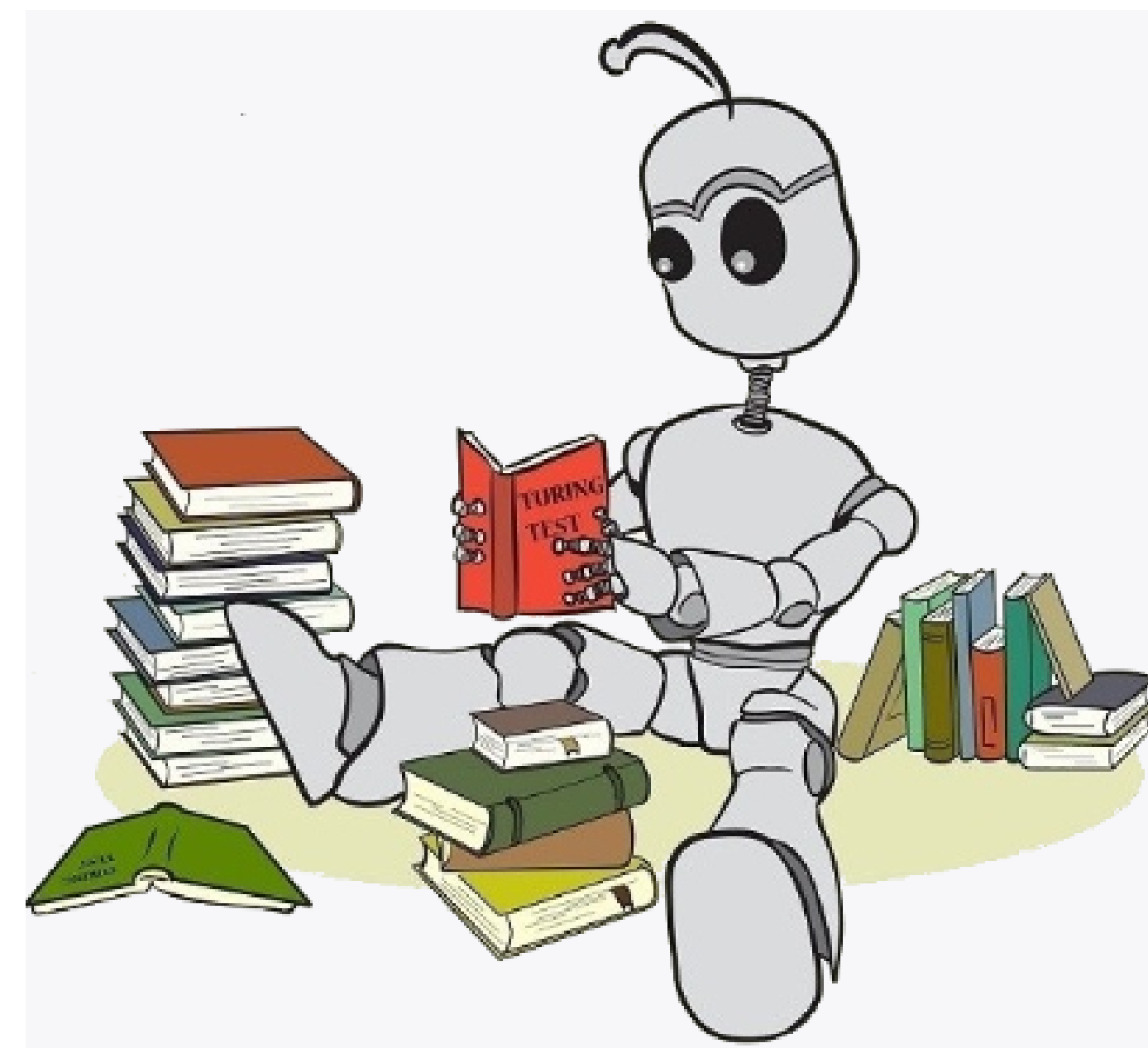
QCRI
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Summary

- Testing with quiz questions helps learning and retention, but creating such questions is a tedious and time-consuming task.
- We built a system that automatically generates multiple-choice questions.
- We fine-tuned T5 models to generate
 - answers
 - answer distractors
 - answer-aware questions
- We trained on open-source datasets.
- We open-sourced our code and training scripts.



System



Leaf

Multiple-Choice Question Generation

Generating multiple-choice questions suitable for knowledge assessment in classroom or industry settings.

Try Demo

Oxygen is the chemical element with the symbol O and atomic number 8. It is a member of the chalcogen group in the periodic table, a highly reactive nonmetal, and an oxidizing agent that readily forms oxides with most elements as well as with other compounds. Oxygen is Earth's most abundant element, and after hydrogen and helium, it is the third-most abundant element in the universe. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O₂. Diatomic oxygen gas currently constitutes 20.95% of the Earth's atmosphere, though this has changed considerably over long periods of time. Oxygen makes up almost half of the Earth's crust in the form of oxides.

Number of questions ...

Generate

Question 1

What group is Oxygen a member of?

Chalcogen group ✓

Oxygen group ✗

Diatomic group ✗

Add answer...

Approve

Question 2

What is the chemical element with the symbol O and atomic number 8?

Oxygen ✓

Dioxygen ✗

Carbon dioxide ✗

Nitrogen ✗

Add answer...

Approve

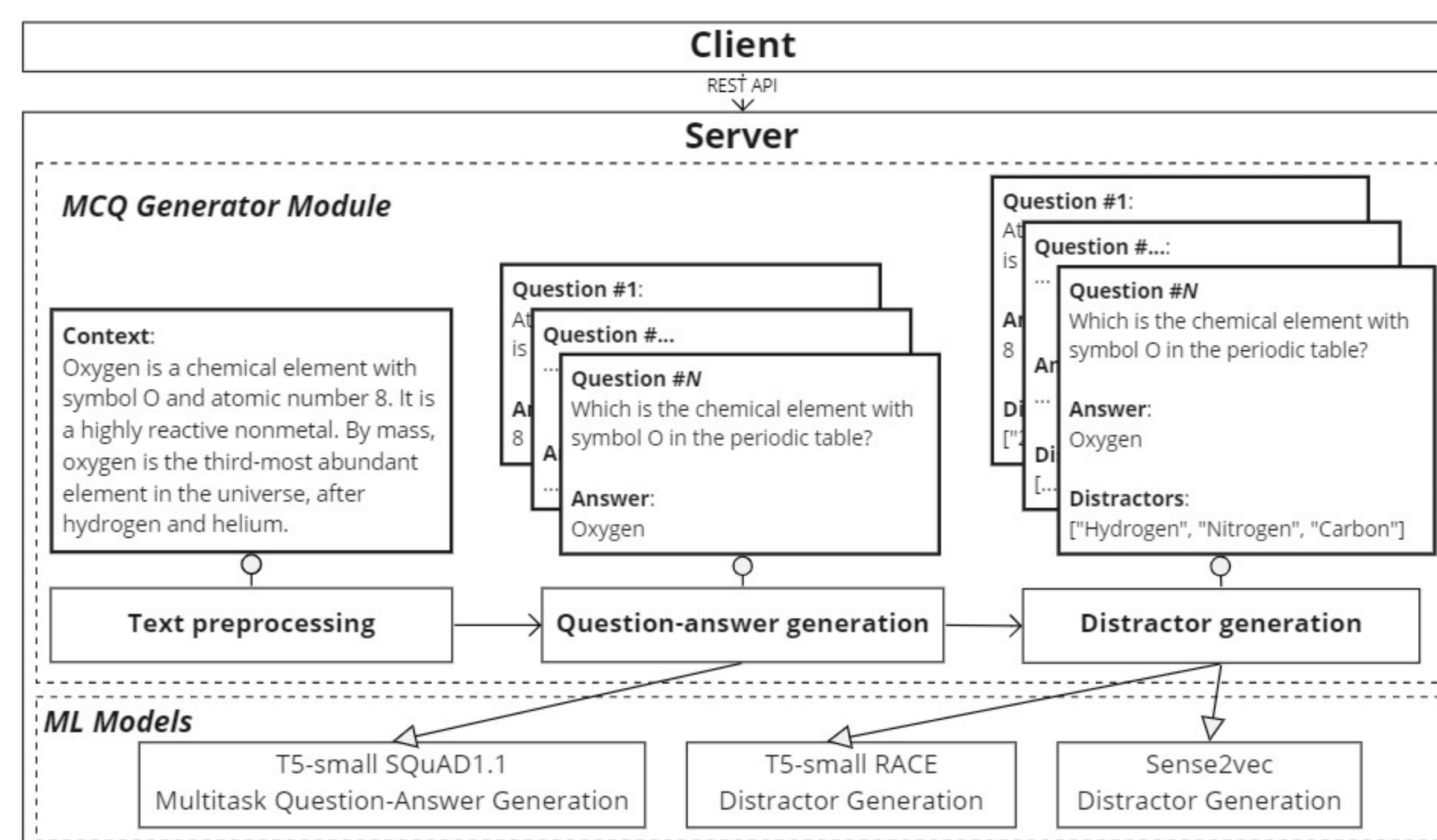
Question Generation

- Based on a fine-tuned T5 Transformer model.
- Using the open-source dataset SQuAD1.1 containing 100,000+ question-answer pairs.
- Multitask model for answer-aware questions generation and question and answer generation from context.

Distractor Generation

- Fine-tuned a version of the T5 Transformer model to generate three distractors.
- Trained on RACE, an open-source dataset containing 100,000 multiple-choice questions.
- Used sense2vec word embeddings for greater variety of distractors.

Architecture



- The instructor can input their educational content and desired number of questions and can then review the generated questions in our web platform.
- After pre-processing the text, the *Multiple Choice Question Generator Module* generates and returns the required number of question-answer pairs and distractors.
- Separating the *ML Models* achieves better abstraction and higher flexibility.

Use Cases

In schools:

- **Students** can self-assess.
- **Instructors** can quickly and efficiently detect knowledge gaps.

In industry:

- New **employees** can be onboarded faster.
- **Managers** can make better strategic decisions.

On other platforms:

- Content for **learning management systems**, which increases user retention.
- Data for training conversational agents or question answering systems.