

AITutor

Team: Darpan Shah, Devanshi Mehta, Dhaval Patel, Ilesh Sharda,
Rishi Singh
Instructor: Chris Asakiewicz

Introduction & Background

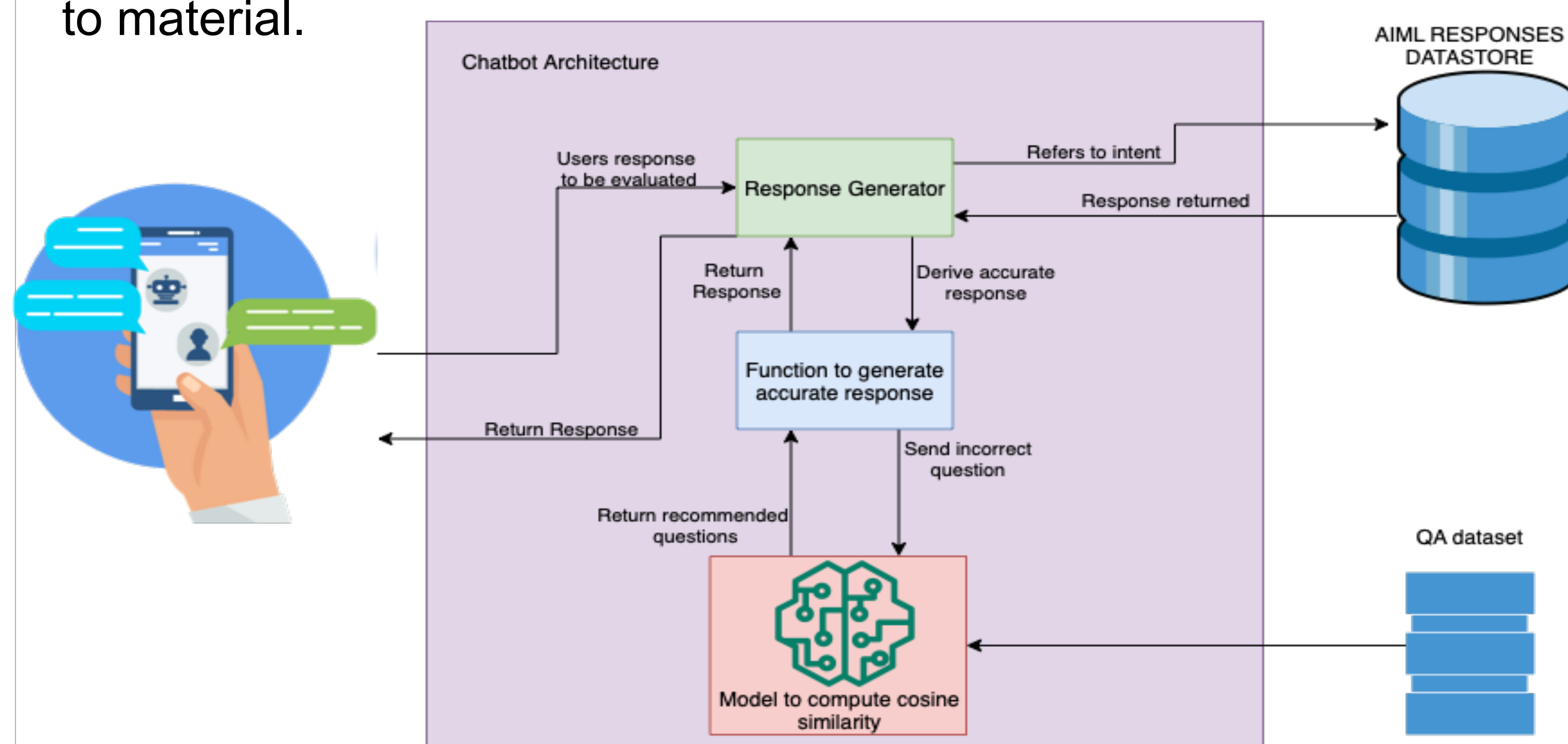
- The emergence of AI-powered education concepts has begun generating new teaching and learning approaches.
- AI system can enhance the collaborative learning experience of students by assisting in answering queries as virtual agents, monitoring the collaboration process, and **identifying a gap in knowledge**.
- AITutor is a chatbot that uses a Multiple-Choice testing format for high school students to identify a gap in their science knowledge through quizzing the student personalizing their learning experience by recommending them questions.

Research Question

- To create a cognitive application that can tailor learning and testing for students
- Identifying and addressing a gap in student's knowledge by assessing them and recommending areas of weakness.
- Measures student's confidence post personalized learning experience to gain an understanding of the success of the application.

Model Architecture and Dataset

The application has been coded using Python and to create the user interface Tkinter has been used to provide a console based design. This design choice is simple and useful and provides the user a straightforward experience. Utilizing the OpenBookQA Dataset which is found on the Allen Institute for AI data repository consists of 5000 questions and six columns with great interest in the question stem and answer key. The application uses Artificial Intelligent Markup Language (AIML) to store the intents for the application and provide accurate responses to the user by making use of tags. The application identifies a gap in the students understanding using incorrect responses to questions and recommends provides questions the student is likely to get incorrect. At the end of the quiz the AITutor recommends areas where the student can brush up their knowledge and provides links to material.



Natural Language Processing

- The Cognitive Application access the student with five random science questions in order to identify a gap in knowledge.
- The application informs students whether the answer is correct or incorrect along with the correct answer and stores the question the student has gotten incorrect.
- Using the **bert-base-nli-mean-tokens** a pretrained model the questions stem are vectorized and stored as a numpy array.
- The pretrained model is known as sentence transformer which is able to map sentences to a 768 dense dimensional vector space.
- Chronologically the question with an incorrect responses has been is used in the cosine similarity check. The top 5 questions with the highest similarity are individually returned to the user.

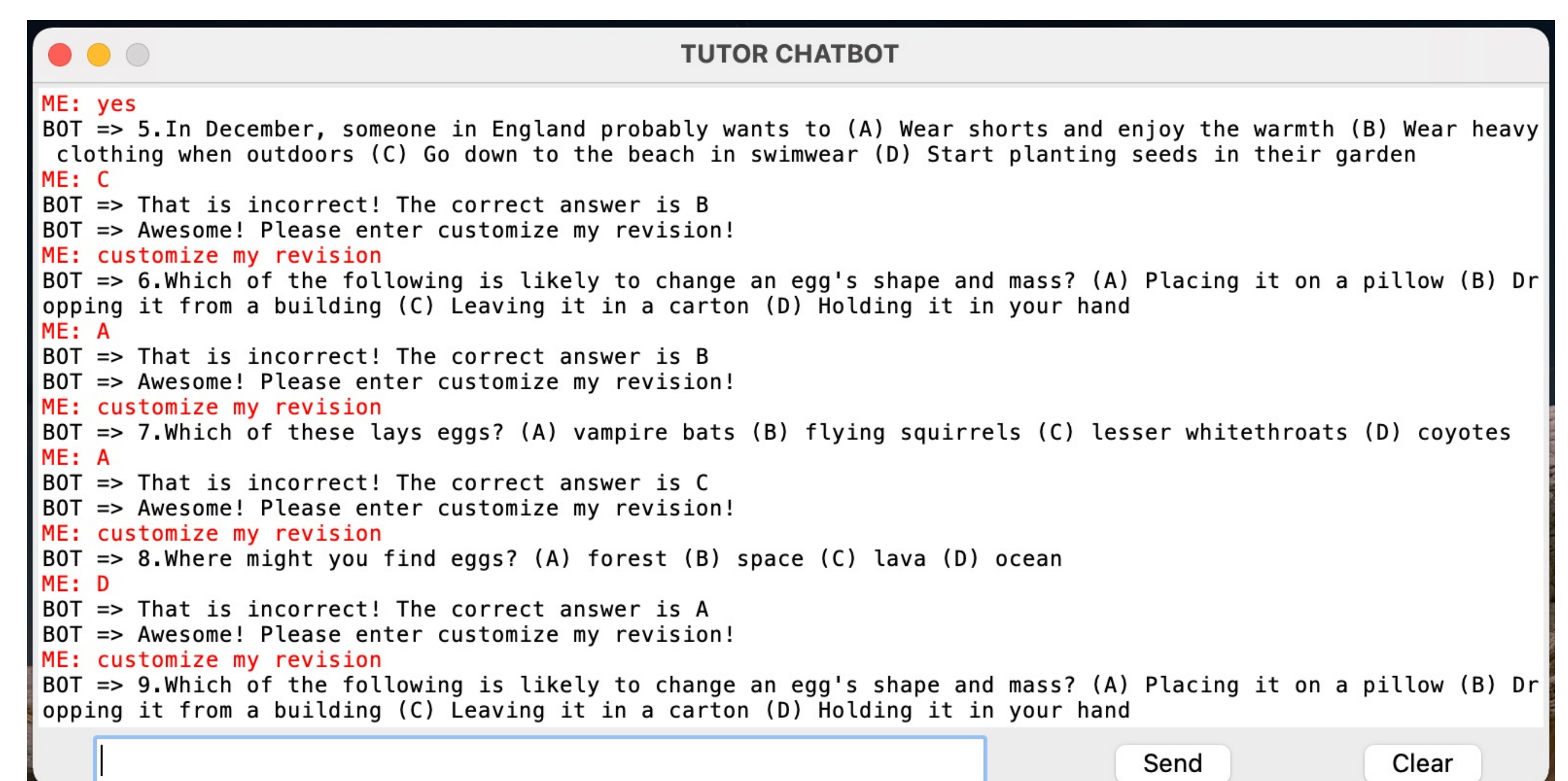
Cosine Similarity Check

- The cosine similarity check is used to retrieve a numeric distance between an a dense vectorized question that has been answered incorrectly by the user and the rest of the questions in the dataset.
- As seen in the figure below the first selected question is 'The sun is responsible for' has highest similarity with Q.107 'the sun is source of...'.
- Using the sklearn cosine similarity function the application computes the cosine similarity distance between an individual question and the rest of dataset along with sorting the questions by similarity value and personalizing the quiz by returning the most similar question.

	Question Stem	similarity
0	The sun is responsible for	1.000000
107	The sun is a source of which, first and foremost?	0.893895
3674	The sun is	0.890997
525	An example of how sunlight is a source of ligh...	0.869584
617	sunlight is a heat source emitted from	0.852023
...
1110	A group of bunnies move into a field and start...	-0.150340
363	After a long winter's hibernation, a female be...	-0.159723
714	A pack of wolves eat the sheep before a grizzl...	-0.168855
21	A mouse is being hunted by a Great Horned Owl...	-0.169972
1495	A squirrel eats all of the acorns in a tree. T...	-0.207894

4957 rows x 2 columns

AITutor UI



In the figure above the tutor chatbot has been able to personalize the quiz dependent on the questions the student has got wrong using pretrained sentence-transformers and cosine similarity. The AITutor is also able to extract keywords from the questions the student has gotten incorrect and return them as topics to the user along with utilizing the WikipediaAPI and returning hyperlinks that will direct the student to information about these topics. To provide the student with personal questions of their wish we have also incorporated a question generation component where the user enters a topic and a fresh new question with multiple choice answers is returned to the user.

Business Impacts

- The business impact of this application from the anticipated results will see a functional chatbot that will be used by students studying science in high school and aid these students with their revision by identifying their gap in knowledge and improve their revision sessions.
- This application should see students gain more confidence in their academics and take their revision and understanding of the science subject into their own hands.

Future Scope

In the future we plan to automate a dynamic question bank that mines data from web and forms MCQ like question generation functionality. This will improve our chatbot by creating fresh content and keeping the dataset up to date. Along with this we aim to improve our topic modelling so more precise recommendations can be returned.