Kwame for Science: An AI Teaching Assistant Based on Sentence-BERT for Science Education in West Africa

George Boateng[1,2], Samuel John[1], Andrew Glago[1], Samuel Boateng[1], and Victor Kumbol[1,3]

[1] SuaCode.ai, Inc., U.S.
[2] ETH Zurich, Switzerland
[3] Charite Berlin, Germany
{jojo, samuel.john, andrew.glago, samuel.boateng, victor}@suacode.ai

Abstract. Africa has a high student-to-teacher ratio which limits students' access to teachers. Consequently, students struggle to get answers to their questions. In this work, we extended Kwame, our previous AI teaching assistant, adapted it for science education, and deployed it as a web app. Kwame for Science answers questions of students based on the Integrated Science subject of the West African Senior Secondary Certificate Examination (WASSCE). Kwame for Science is a Sentence-BERT-based question-answering web app that displays 3 paragraphs as answers along with a confidence score in response to science questions. Additionally, it displays the top 5 related past exam questions and their answers in addition to the 3 paragraphs. Our preliminary evaluation of the Kwame for Science with a 2.5-week real-world deployment showed a top 3 accuracy of 87.5% (n=56) with 190 users across 11 countries. Kwame for Science will enable the delivery of scalable, cost-effective, and quality remote education to millions of people across Africa. [4]

Keywords: Virtual Teaching Assistant · Educational Question Answering · Science Education · NLP · BERT · SBERT · West Africa

1 Introduction

The COVID-19 pandemic has exacerbated the already poor educational experiences of millions of students in Africa who were grappling with educational challenges like poor access to computers, the internet, and teachers. In 2018, the average student-teacher ratio in Sub-Saharan Africa was 35:1 which is higher compared to 14:1 in Europe [10]. In this context, students struggle to get answers to their questions. Hence, offering quick and accurate answers, outside of the classroom, could improve their overall learning experience. However, it is difficult to scale this support with human teachers.

In 2020, we developed Kwame [3], a bilingual AI teaching assistant that provides answers to students' coding questions in English and French for SuaCode, a smartphone-based online coding course [5,4]. Kwame is a deep learning-based question answering system that finds the paragraph most semantically similar to the question via cosine similarity with a Sentence-BERT model. We extended Kwame to work for science education and deployed it as a web app. Specifically, Kwame for Science [5] answers questions of students based on the Integrated Science subject of the West African Senior Secondary Certificate Examination (WASSCE). This is a core subject that covers various aspects of science such as biology, chemistry, physics, earth science, and agricultural science. It is mandatory for senior high school students in the West African Education Council (WAEC) member countries (Ghana, Nigeria, Sierra Leone, Liberia, and The Gambia).

There are virtual teaching assistants (TA) such as Jill Watson [7,6], Rexy [1], and a physics course TA [11] and Curio SmartChat (for K-12 science) [9] (see [2] for a detailed description of related work). These works are focused on answering logistical questions, except Curio SmartChat. In comparison to Curio SmartChat which is the closest work to ours, our work uses a state-of-the-art language model (Sentence-BERT) relative to theirs (Universal Sentence Encoder). Also, our work is the first to be developed and deployed in the context of high school science education in West Africa.

2   Kwame for Science System Architecture

Kwame for Science is a Sentence-BERT-based question-answering web app that displays 3 paragraphs as answers along with a confidence score which represents the similarity score in response to science questions (Figure 1). Additionally, it displays the top 5 related past exam questions and their answers in addition to the 3 paragraphs. We used a Sentence-BERT (SBERT) model that was pre-trained on a large and diverse set of question-answer pairs. We used the SBERT model as it was, with plans for fine-tuning after real-world data collection especially since exploratory evaluation for our science use case showed it had decent performance.

When a user types a question in the web app, our system computes an embedding of the question using the SBERT model. Next, it computes cosine similarity scores with a bank of answers (which are paragraphs from our knowledge source), retrieves, and returns the top 3 answers along with a confidence score and any figures or images referenced in that paragraph to the web app. Additionally, it computes cosine similarity scores with a bank of past exam questions, retrieves, and returns the top 5 related questions and their answers, along with confidence scores. The web app then displays the answers and the related past exam questions that are above a preset similarity score threshold. If no answer is above the threshold, a message is shown saying the question could not be answered using the knowledge source of that subject. We precomputed embeddings for fast

---

[5] http://kwame.ai/

real-time retrieval and saved them as indices in ElasticSearch which we hosted on Google Cloud Platform.
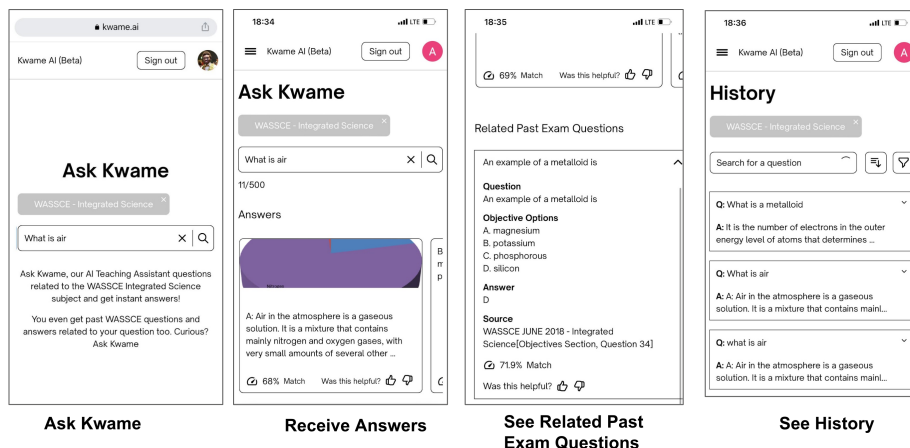

Fig. 1. Screenshots of Kwame for Science

3   Dataset Curation and Preprocessing

Given that our goal was for Kwame to provide answers based on the Integrated Science subject of the WASSCE exam, our training data and knowledge source had to cover the topics in the WASSCE Integrated Science curriculum. We sought to use one of the approved textbooks in Ghana. Unfortunately, their copyrights did not permit such use and the publishers were unwilling to partner with us. Consequently, we searched for free and open-source books and datasets that fulfilled our needs. We came across a middle school science dataset — Textbook Questions Answering (TQA) [8] which was curated from the free and open-source textbook, CK-12. Our exploration of the dataset revealed that though it covered several of the WASSCE Integrated Science topics, it lacked others, particularly those related to agricultural science. Consequently, we additionally used a dataset based on Simple Wikipedia to cover those gaps. We used Simple Wikipedia since its explanations were simple and better suited for middle school and high school students compared to regular Wikipedia.

We parsed the JSON files of the dataset into paragraphs. We also extracted figures that were referenced in the paragraphs so they could be returned to students along with the answers. We then split the paragraphs into groups of 3 sentences, computed embeddings, and indexed them using ElasticSearch to enable fast retrieval and run time. These constituted the answers returned for questions. Furthermore, we augmented our question-answering with curriculum-specific content. In particular, we created question-answer pairs using WASSCE

questions that cover exams from 2000 to 2020. The exam has three parts, objectives (multiple-choice), theory, and practicals. Similar to the paragraphs, we computed embeddings of the questions and indexed them using ElasticSearch. These constituted the related past questions (with answers) returned when a question is asked.

4   Preliminary Evaluation and Results

We launched the web app in beta on 10th June 2022. Users could provide feedback by upvoting or downvoting answers in response to the question "Was this helpful?." To evaluate Kwame for Science, we used the metrics top 1 and top 3 accuracies. Top 1 accuracy quantifies performance assuming only one answer was returned and voted on. Top 3 accuracy refers to the performance where for each question that received a vote, at least one answer was rated as helpful out of the 3 answers that were returned. The statistics for the deployment between 10th June 2022 and 27th June 2022 (2.5 weeks) are 190 users across 11 countries (6 in Africa), 433 questions with the metrics 71.8% top 1 accuracy (n=117 answers), and 87.5% top 3 accuracy (n=56 questions). The top 3 accuracy result is good, showing that Kwame for Science has a high chance of giving at least one useful answer among the 3. Some challenging cases occurred when there were typos in the spelling of scientific words and the questions were related to topics outside the scope of the knowledge source. Also, some unhelpful answers were cases where the returned paragraph was incomplete due to issues with the dataset.

5   Conclusion

In this work, we developed and evaluated Kwame for Science which provides instant answers to the Science questions of students across West Africa. Our future work will fine-tune the SBERT model using the real-world votes on answers to improve its accuracy. Also, we will make Kwame for Science available in local languages across Africa, and available via offline channels such as SMS, USSD, and toll-free calling. Kwame for Science will enable the delivery of scalable, cost-effective, and quality remote education to millions of people across Africa.

6   Acknowledgement

References

1. Benedetto, L., Cremonesi, P.: Rexy, a configurable application for building virtual teaching assistants. In: IFIP Conference on Human-Computer Interaction. pp. 233–241. Springer (2019)

2. Boateng, G.: Kwame: A bilingual ai teaching assistant for online suacode courses. arXiv preprint arXiv:2010.11387 (2020)
3. Boateng, G.: Kwame: a bilingual ai teaching assistant for online suacode courses. In: International Conference on Artificial Intelligence in Education. pp. 93–97. Springer (2021)
4. Boateng, G., Annor, P.S., Kumbol, V.W.A.: Suacode africa: Teaching coding online to africans using smartphones. In: Proceedings of the 10th Computer Science Education Research Conference. pp. 14–20 (2021)
5. Boateng, G., Kumbol, V.W.A., Annor, P.S.: Keep calm and code on your phone: A pilot of suacode, an online smartphone-based coding course. In: Proceedings of the 8th Computer Science Education Research Conference. pp. 9–14 (2019)
6. Goel, A.: Ai-powered learning: Making education accessible, affordable, and achievable. arXiv preprint arXiv:2006.01908 (2020)
7. Goel, A.K., Polepeddi, L.: Jill watson: A virtual teaching assistant for online education. Tech. rep., Georgia Institute of Technology (2016)
8. Kembhavi, A., Seo, M., Schwenk, D., Choi, J., Farhadi, A., Hajishirzi, H.: Are you smarter than a sixth grader? textbook question answering for multimodal machine comprehension. In: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition. pp. 4999–5007 (2017)
9. Raamadhurai, S., Baker, R., Poduval, V.: Curio smartchat: a system for natural language question answering for self-paced k-12 learning. In: Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. pp. 336–342 (2019)
10. Unesco. pupil-teacher ratio sub-saharan africa. https://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS?locations=ZG (Feb 2020)
11. Zylich, B., Viola, A., Toggerson, B., Al-Hariri, L., Lan, A.: Exploring automated question answering methods for teaching assistance. In: International Conference on Artificial Intelligence in Education. pp. 610–622. Springer (2020)