2024 International Conference on Intelligent and Innovative Computing Applications (ICONIC)
7th-9th November 2024, Pearle Beach Resort & Spa, Flic en Flac, Mauritius
https://doi.org/10.59200/ICONIC.2024.003

AI-Enhanced STEM Education in Sub-Saharan Africa and Indian Ocean Island Nations using Whatsapp and GPT-4*x* in Local Languages

Laurie Butgereit
Nelson Mandela University
Gqeberha, South Africa
laurie.butgereit@gmail.com

Abstract
In a 2016 report, the World Bank warned that Sub-Saharan African countries needed to boost performance in STEM subjects in order to become competitive in the global marketplace. The report presented the World Bank's fears that Sub-Saharan Africa's economic development is being impeded by the lack of high-quality education especially in STEM subjects. In particular, poor performance in mathematics at both the primary and secondary school levels created barriers to improved economic outcomes. This paper looks at how the popular WhatsApp chat mobile app (along with web interfaces when WhatsApp is not appropriate) can be linked to OpenAI's GPT-4 artificial intelligence system to provide tutoring in STEM subjects in a variety of human languages used in Sub-Saharan Africa.

Keywords: GPT-4x, WhatsApp, tutoring, STEM

1. Introduction

Education in STEM (Science, Technology, Engineering and Mathematics) and non-STEM subjects varies in a number of key areas. In general, STEM education focuses on specialised and practical skills such as mathematics, computer science, astronomy, and engineering. Non-STEM education, on the other hand, covers a wider range of subjects including history, sociology, philosophy, music, art, literature, and languages. The methods that are used to teach the two types of subjects also vary. The education of STEM subjects usually relies on solving practical problems, laboratory exercises, and direct application of skills. The education of non-STEM subjects, on the other hand, often relies on theoretical subjects, exploring concepts, and discussion groups. STEM education emphasises hard skills, industry-specific knowledge, and technical proficiency. Non-STEM education emphasises communication, emotional intelligence, and critical thinking.

In a STEM subject, asking a student a question such as "What is the area of a farmer's field which is 100 meters long and 250 meters wide?" has a very specific right answer. In a non-STEM subject, on the other hand, asking a student a question such as "What was the cause of the First World War?" does not have a specific right answer and could be used as an introduction to a wide ranging discussion. A recent World Bank study found that Sub-Saharan Africa needs to boost performance in STEM in order to become competitive in the global market. This is especially true due to the fact that the global market being driven by new technologies (Bethell, 2016). This paper specifically looks at how artificial intelligence (AI) can be used to assist with STEM education. This is not to say that AI cannot be used in non-STEM education. This paper, however, only investigates the use of AI in STEM education.

The artifact described in this paper, TutorBot, links WhatsApp (in addition to a web interface) to the popular GPT-4*x* API to provide an AI-enhanced academic tutor in a number of STEM subjects. In the expression GPT-4*x*, the *x* is used to denote a number of sub-releases of GPT-4 and this paper will not generally distinguish between them unless necessary. The TutorBot platform has been already used in a number of academic projects which have already been reported and those projects will be simply summarised in this paper. The platform is also being used in a number of ongoing commercial tutoring projects which will not be reported in this paper since the intellectual property of those commercial project is not in the public domain. These commercial

tutoring projects, however, were all conducted in English. This paper specifically looks at research projects where TutorBot had been configured to operate in a number of languages used in Sub-Saharan Africa and in the Indian Ocean Island nations including Afrikaans, Amharic, Arabic, isiZulu, and both Mauritian Creole and Seychellois Creole.

Section 2 provides background information on AI, GPT-4*x*, chatGPT, and OpenAI. It is important to note that this is not a paper about artificial intelligence *per se*. This paper is about using artificial intelligence in STEM education. The background information in this paper about artificial intelligence is limited. Section 3 provides information on the methodology of developing the TutorBot artifact along with the different methodologies of a number of projects which used TutorBot in languages used in Sub-Saharan Africa. In addition, there is a brief discussion of the ethics on these various projects in Section 3. Although the TutorBot artifact has been reported previously, a brief description of the artifact is provided in Section 4. Interesting results of the various projects are provided in Section 5. The business benefits (or lack thereof) are provided in Section 6. Concluding remarks can be found in Section 7.

2. Background

The concept of artificially intelligent objects or beings has been present in literature and mythology for millenia. In Greek mythology, Pygmalion made an ivory statue which represented his ideal concept of woman. In the myth, Pygmalion subsequently falls in love with the statue and the statue is brought to life by the Goddess Venus (Gorlinkski, 1998). In the 1800s, Mary Shelley penned the story of Dr Frankenstein who creates a being and animates it using electricity instead of using a goddess (Byrd & Paquette, 2023; Shelley, 1816).

In 1950, the famous Alan Turing grappled with the question of whether or not machines could *think*. In his paper, he proposed a game he called *The Imitation Game* where a person could ask questions of two different entities. In this game, one of the entities was another human being and one of the entities was a computer. By asking questions and receiving responses, the person needed to determine which entity was the other human being and which entity was the computer (Turing, 1950).

The first actual use of the expression *artificial intelligence*, however, came in 1955 when McCarthy *et al* proposed "that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College..." The study would attempt to determine how "to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans…" The estimated cost of this study (including salaries and travel expenses) was US$13,500 (McCarthy et al., 1955).

Fast forward to 2022 when a research center named OpenAI released a website called chatGPT (OpenAI, 2022) which could answer a question remarkably like a human would answer the question. The API behind chatGPT is known as GPT with various versions numbers appended to the name. At the time of writing this paper, the latest version is GPT-4o (*o* standing for Omni). For the scope of this paper, however, the various sub-version of GPT-4 will be simply called GPT-4*x*. The technical report on GPT cites remarkable performance on many standardised tests (OpenAI, 2023). With such remarkable performance, it is clear that GPT could be used in certain types of educational artifacts.

3. Methodology and Ethics

This section describes two different types of projects: 1) the development of the generalised TutorBot 2) the development of various AI-enhanced language based STEM tutors. In addition there is a brief discussion of the ethics involved.

The development of the TutorBot artifact followed an Agile Methodology (Alliance, 2017; Fowler & Highsmith, 2001) common with most modern commercial software development. The methodology emphasizes frequent delivery of working software over short time frames. Although Agile does emphasize face-to-face collaboration, the collaboration in this particular case happened over video conferencing systems. A technical description of the TutorBot artifact can be found in Section 4. The ownership of the intellectual property of the TutorBot artifact is described in the Acknowledgements section.

The TutorBot artifact allows for easy configuration of specific subjects to tutor and easy configuration of the human languages to be used in the tutoring sessions. These configurations of individual projects used Design

Science Research and involved technical people, educators, and prompt engineers.

Design Science Research is used to solve real problems in the real world (Hevner & Chatterjee, 2010). Design Science Research supports a three-cycle view (Hevner, 2007) of the research project which ensures that the project solves a real problem. There is a Rigor Cycle which ensures that the project uses scientifically rigorous methods. There is a Relevance Cycle which ensures that the project solves a real problem. And there is a Design Cycle which links everything together.

The development of the TutorBot artifact was strictly a commercial software development project and did not require any ethics approval. A number of the individual tutoring instances did require ethics approval. In such cases, a full ethics application was submitted to the educational institution involved and an approval was received. These tutoring instances were deployed with real students in real situations and, therefore, needed ethics approval. The conversational text recordings of the participating students was reviewed. The participating students filled in questionnaires and often had face-to-face interviews with the researchers.

A number of the individual tutoring instances were not deployed to students but were created and evaluated by collaborating researchers or by linguists hired by the researchers. In such cases, since no experiments were being done on any human subjects, no ethics approvals were required. A number of the individual tutoring instances are operational for commercial entities. These types of instances are not research projects and no ethics approvals are required. It is important to also note that at the time of writing this paper, all of these commercial deployments operated in English.

In Section 5 (Results section), some of the examples were taken from log files of participating students where the project had a full ethics approval from the educational institution involved in that project. Other examples in other languages were generated by either collaborators or hired linguists and did not require any ethics approval.

4. Technology Description

The TutorBot artifact used in the various projects described in this paper has two user interfaces. One of the user interfaces is Whatsapp and the other user interface is a typical web interface presented in a browser. In both cases, the user can pose a question about some STEM subject (such as mathematics or Java programming). The question is routed to the proprietary TutorBot server which checks permissions, logs the message, wraps the message in appropriate prompts and forwards it to GPT-4*x*. When the reply is received from GPT-4*x*, the reply is then logged and forwarded back to the user. (Note that the TutorBot platform was developed before OpenAI released *assistants* and *persistent threads*. The platform may be updated to take advantage of those features in the future.)

There are two different user interfaces which are available and are configured depending on the subject being tutored. For example, in the case of a mathematics tutoring bot, the Whatsapp interface is typically used because the student usually only has his or her cell phone available. In the case of a programming tutoring instance such as Java, then the web interface can be configured since if a student is learning Java Programming, then he or she probably has a laptop or work station available for use.

The TutorBot artifact allows administrators to configure welcome messages and help messages. In addition, prompts to GPT-4*x* can be defined which configure the tutoring instance to tutor specific subjects (such as mathematics or C++ programming) and help define which human languages (such as Mauritian Creole or English) are used in the tutoring sessions. In addition (because of cost issues with GPT-4*x* which is discussed in more detail in Section 6), the size of the conversation context can be configured per project. One of the most important configuration features was to instruct GPT-4*x* *how much* of an answer should be provided to the student or pupil. Should GPT-4*x* just answer the question posed by the student or pupil? Of course not! That would defeat the purpose of a tutor. Configuration options exist to describe how extensive the answers should be, etc.

5. Results

As mentioned previously, there have been a number of research projects using TutorBot which had full ethics approval. These projects included AI-enhanced tutors communicating in English (Seeam & Butgereit, 2024), Arabic (Butgereit et al., 2023), and Sesotho (Smit & Butgereit, 2024). There were additional projects which used language experts and domain experts to evaluate the AI-enhanced tutors and, therefore, did not require ethics approvals.

These projects included tutors communicating in Afrikaans (Butgereit & van Staden, 2023), Amharic (Butgereit & Egu, 2023), Mauritian Creole (Butgereit, Gooroochurn, et al., 2024), Seychellois Creole (Butgereit, Pejakovic, et al., 2024), and isiZulu (Butgereit, 2023). The reader is referred to the specific papers for language specific information. Examples of some of those language projects can be seen in Figures **Error! Reference source not found.**-**Error! Reference source not found.**.

Figure 1: Seychelles Creole example

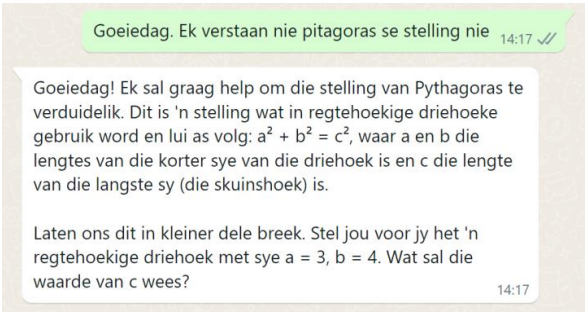

Figure 2: Afrikaans example

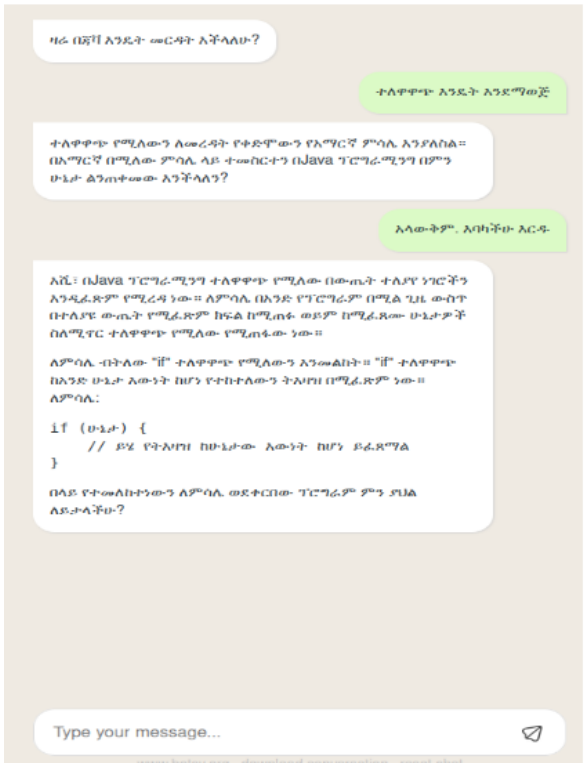

Figure 3: Amharic example

Figure 4: Zulu example

As mentioned before, this paper does not dwell on the various versions of GPT-4*x* which exist. However, in this Results section it is important to clarify this issue. At the time of writing this paper, there are three different versions of GPT-4*x*. These are GPT-4, GPT-4-turbo, and GPT-4o (Omni). All of the research cited at the beginning of this section was done with GPT-4. In summarising the various findings, with the GPT-4 API, there were some mistakes with the non-English languages. Users who were middle-school aged and above could easily understand the mistakes and could still find the AI-enhanced tutor of great benefit to their studies.

OpenAI's press release announcing GPT-4o (Omni), however, promises "significant improvement on text in non-English languages" (OpenAI, 2024). Where possible and where of interest, the authors are undertaking more formal testing of Sub-Saharan African languages and languages used in the Indian Ocean Island nations.

6. Business Benefits

At first glance, one would assume that the use of GPT-4*x* in tutoring would be a net positive for education in Sub-Saharan Africa and the Indian Ocean Island nations. This would be especially useful in rural areas where there might not be tutors or teachers to handle STEM subjects. However, the use of the GPT-4*x* API is not free. At the time of writing this paper even the well know Khan Academy is asking for donations of US$9.00 per month or US$99.00 per year for their GPT-4 enabled tutor (Khan Labs, n.d.). The financial press reports that it costs OpenAI approximately US$700,000 (seven hundred thousand US Dollars) *per day* to run chatGPT (Mok, 2023). OpenAI needs to recoup those costs somehow.

Although OpenAI does openly publish the prices for using the various versions of the GPT APIs (OpenAI, n.d.), it is often hard for researchers to initially estimate the costs that a student will incur on a new project. One of the configurable options of TutorBot is how large of a conversation context must be sent to the GPT-4*x* API. As mentioned before, the TutorBot artifact is also used in commercial installations. The organisation holding the intellectual property of the TutorBot has experience on how to price the products that they offer. That information is beyond the scope of this paper.

Although the process of configuring GPT-4*x* to be a tutor and linking students and pupils to GPT-4 through either Whatsapp or a web interface worked well in many Sub-Saharan African languages and also worked reasonably well in two Creoles from Indian Ocean Island nations funding would need to be obtained in order to roll this out on a wide scale.

7. Conclusion

This paper investigates the use of the popular WhatsApp mobile chat app (and an additional web interface) to link pupils and students to GPT-4*x* to provide tutoring in STEM subjects. For languages which are well resourced, the process of tutoring STEM subjects in those languages works well. For lessor resourced languages, there can be language issues. It is presumed that this is due to a dearth of training text available in those languages when OpenAI was training their large language model. This problem may be alleviated in future releases for the GPT API such as GPT-4o (Omni). Costs are a major issue. The GPT-4*x* library is not free to use. Depending on the type of student or pupil being tutored and the subject matter, gross costs per project can vary. Having said that, in underserved areas where teachers and tutors in STEM subjects are rare or not available, linking WhatsApp to GPT-4*x* to answer pupils' and students' questions in local languages works well.

Acknowledgements

This research was done in conjunction with Herman Martinus of Deep Tutor, a startup in Cape Town, South Africa. The intellectual property of the TutorBot artifact is owned by Deep Tutor. Deep Tutor focuses on the integration of artificial intelligence and education to create innovative learning experiences. They have developed a suite of AI-driven products that cater to diverse educational needs.

8. References

Alliance, A. (2017). *Agile Alliance Website* (Vol. 2018). https://www.agilealliance.org/

Bethell, G. (2016). *Mathematics Education in Sub-Saharan Africa: Status, Challenges, and Opportunities*. The World Bank. https://openknowledge.worldbank.org/entities/publication/0189b507-44cb-5b0b-afa7-bebbb9fc7713

Butgereit, L. (2023, December 6). *Prof Pi: GPT-4 Tutoring Mathematics in African*

Languages. 4th International Conference on Teaching, Assessment and Learning in the Digital Age, Cape Town.

Butgereit, L., & Egu, A. G. (2023). *Using GPT-4 to Tutor Java Programming in Amharic*. First International Conference on the Advancement of Artificial Intelligence in African Context 2023, Arusha, Tanzania.

Butgereit, L., Gooroochurn, M., & Bosquet, Y. (2024). *Dispensing Climate Change Information in Mother Tongue using a GPT-4 Based a ChatBot: A Case with Mauritian Creole*. 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering, Maldives.

Butgereit, L., Martinus, H., & Muna Mahmoud, A. (2023). Prof Pi: Tutoring Mathematics in Arabic Language using GPT-4 and Whatsapp. *Proceedings INES 2023*, 161–164.

Butgereit, L., Pejakovic, C., & Athanase, J.-M. (2024). *Using AI to Support Young Mathematics Learners in Seychellois Creole*. Digital2k, Mauritius.

Butgereit, L., & van Staden, A. (2023, October). *An AI Contribution to African Economic Development through Mother Tongue Education—A Linguistic Evaluation of an African-based AI Mathematics Tutoring Bot*. IEEE 5th International Conference on ICT for Development for Africa, Ethiopia.

Byrd, J., & Paquette, P. (2023). Frankenstein: A Creation of Artificial Intelligence? *AI & Society*, *38*, 1–12. https://doi.org/10.1007/s00146-021-01298-7

Fowler, M., & Highsmith, J. (2001). The agile manifesto. *Software Development*, *9*(8), 28–35.

Gorlinkski, V. (1998). Pygmalion. In *Britannica*. https://www.britannica.com/topic/Pygmalion

Hevner, A. R. (2007). The Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Hevner, A. R., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice* (Vol. 22). Springer Verlag.

Khan Labs. (n.d.). *Khan Labs Webpage*. Retrieved October 3, 2023, from https://www.khanacademy.org/khan-labs#khanmigo

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955, August 31). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

Mok, A. (2023, April 20). ChatGPT could cost over $700,00 per day to operate. *Business Insider*. https://www.businessinsider.com/how-much-chatgpt-costs-openai-to-run-estimate-report-2023-4

OpenAI. (n.d.). *Pricing*. OpenAI Pricing Page. https://openai.com/pricing

OpenAI. (2022, November 30). *Introducing chatGPT*. https://openai.com/blog/chatgpt

OpenAI. (2024, May 13). *Hello GPT-4o*. https://openai.com/index/hello-gpt-4o/

OpenAI, O. (2023). *GPT-4 Technical Report*. https://arxiv.org/abs/2303.08774

Seeam, A., & Butgereit, L. (2024). *University Students' Perceptions of AI Programming Tutors: A Chatbot Integration Study in Programming Education*. Paper under review for Digital2k, Mauritius.

Shelley, M. W. (1816). *Frankenstein; or, the Modern Prometheus*. https://www.gutenberg.org/files/84/84-h/84-h.htm

Smit, I., & Butgereit, L. (2024). *The value of a Sesotho bot as tutor for an introductory Python programming class*. Digital2k 2024, Mauritius.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, A Quarterly Review of Psychology and Philosophy*, *LIX*(236).