International Journal of Social Robotics (2026) 18:62
https://doi.org/10.1007/s12369-026-01397-9

A Tale of Two Tutors: Swahili-Speaking Robot Tutoring After Classroom Math Lessons in Tanzanian Primary Schools

Edger P. Rutatola[1,2] · Qiaoqiao Ren[1] · Maria Jose Pinto-Bernal[1] · Koen Stroeken[3] · Tony Belpaeme[1]

Received: 15 June 2024 / Revised: 18 April 2026 / Accepted: 23 April 2026 / Published online: 18 May 2026

Abstract

This study explores the potential and integration of a Swahili-speaking social robot as a mathematics tutor in Tanzanian primary schools. Leveraging the GPT-3.5 Large Language Model (LLM) and a NAO social robot, the research investigates the feasibility and effectiveness of using state-of-the-art technologies to support mathematics education in a low-resource setting. Because classroom lessons in such contexts often provide limited opportunities for individualised support, the robot was designed to serve as an after-class remedial tutor following a lesson on the same topic taught by a human teacher. Through a series of sessions conducted in five public primary schools involving 26 third- and fourth-grade pupils, the study assessed pupils' cognitive learning outcomes and their perceptions of the social robot as a tutor. The robot's tutoring approach incorporated contextual information and pupils' progress to provide personalised feedback. The results show a significant improvement in pupils' understanding of the mathematics topic following the robot's tutoring, supported by qualitative feedback highlighting the robot's friendliness, knowledgeability, and effective tutoring methods. The study underscores the potential of AI-powered social robots to complement classroom teaching in resource-constrained environments and offers insights for future research and development in this domain.

Keywords Large language models · Social robots in education · Mathematics education · GPT models · Swahili · Low-resource

1 Introduction

1.1 Mathematics Education in Tanzania

Science and technology are at the core of the development innovations we experience today, fuelling economic growth and improving the livelihoods of many nations. It is also undisputed that mathematics forms the foundation for understanding scientific and technological concepts, making its proficiency essential in the modern world [1, 2]. However, pupils in the United Republic of Tanzania still struggle with mathematics at all levels, from primary schools to universities [3, 4]. The country faces significant educational system challenges such as large class sizes and unfavourable teacher-pupil ratios, which make the classes unmanageable [3, 5]. The situation is more critical in resource-limited rural areas, where the teacher-pupil ratio is as extreme as 1:133 [6]. These limitations make it difficult for teachers to provide effective individualised instructions and personalised feedback to each pupil, thus impeding pupils' performance in the subject. For example, in the 2022 and 2023 national

✉ Edger P. Rutatola
edger.rutatola@ugent.be

Qiaoqiao Ren
qiaoqiao.ren@ugent.be

Maria Jose Pinto-Bernal
mariajose.pintobernal@ugent.be

Koen Stroeken
koen.stroeken@ugent.be

Tony Belpaeme
tony.belpaeme@ugent.be

1 IDLab-AIRO, Ghent University – imec, Ghent, Belgium

2 Department of Computing Science Studies, Faculty of Science and Technology, Mzumbe University, Morogoro, Tanzania

3 Centre for Anthropological Research on Affect and Materiality (CARAM), Ghent University, Ghent, Belgium

primary school leaving examinations, 46.72% and 51.17% of pupils failed mathematics, respectively [7].

1.2 Large Language Models and Social Robots

Since the advent of Large Language Models (LLMs) such as OpenAI's (Chat-)GPT in late 2022, generative artificial intelligence (AI) has advanced significantly, driven by their remarkable breakthroughs and democratic accessibility. These models have opened doors to new realms in the development of interactive and engaging applications that would otherwise be out of reach of many. LLMs enable creating interactive and engaging applications with remarkably fluent, eloquent, and plausible human-like dialogue capabilities [8]. A study by Noy and Zhang has shown their effectiveness in boosting productivity across various industries [9].

The education sector is also witnessing a transformation driven by LLMs. These models demonstrate an exceptional ability to generate high-quality content that can match that of experts, even worthy of publishing in reputable journals [10]. Researchers are actively leveraging this potential by integrating LLMs into educational tools. For example, a study by Verheist et al. [11] successfully leverages GPT-3.5 to help children learn a second language without any assistance from a human domain expert. Similar positive cognitive and affective outcomes have been observed in other studies exploring LLMs for various subjects and demographics [12, 13].

Despite their undeniable benefits, LLMs are virtual, limiting the user experience in the physical dimension. This is where social robots bridge the gap, as they offer a physical presence that embodies the virtual content of LLMs leading to a manifestation of a physical tutor. A review of various applications of social robots in education shows that the physical presence offered by social robots significantly enhances children's learning [14]. Therefore, exploiting the synergy of LLMs and social robots is essential for maximising the impact of the two.

1.3 Research Questions

Considering the challenges facing mathematics teaching and learning in primary schools in Tanzania and the advances in LLMs, this study explores the potential of developing a Swahili-speaking social robot powered by an LLM. Swahili, Tanzania's national language and primary medium of instruction in public primary schools, is considered a low-resource language with limited training data for LLMs [15]. Specifically, in CommonCrawl –which provides the training data for GPT models – Swahili makes up only 0.0080% of training data [16]. Therefore, we aim to assess whether and how the current state of LLMs (particularly GPT-3.5, the LLM under study here) can generate sufficient learning content in Swahili to achieve the anticipated educational impact. This will provide valuable insights to researchers in similar low-resource contexts. Thus, this study seeks to answer the following research questions:

i. How can LLMs be adapted to generate reliable Swahili-language tutoring interactions for a social robot in Tanzanian primary schools?
ii. Does tutoring delivered by an LLM-powered social robot after classroom instruction lead to improvements in pupils' understanding of a mathematics topic?
iii. How do Tanzanian primary school pupils perceive a social robot as their mathematics tutor?

2 Methods

2.1 Study Context

The study was conducted in the United Republic of Tanzania, involving five public primary schools. We purposely chose to run our study in public (government-funded) schools rather than private schools due to their aforementioned critical conditions, which offered an ideal test ground for piloting a social robot as a tutor. Moreover, 89.2% of primary schools in the country are public schools, with private schools constituting the minority [17]. Two schools were in the Dar es Salaam region, the commercial metropole of Tanzania, representing a best-case scenario environment for the experiment and prospects. The remaining three schools were in the Morogoro region, which has a mix of urban and rural settings [18], offering a diverse perspective to the study in terms of reception and prospects of our intervention. In Dar es Salaam, both schools were in the Ubungo district (urban), while in Morogoro, one school was in the Morogoro Municipal (urban) and the other two were in the Mvomero district (rural).

2.2 Subject and Topic of Intervention

The social robot was customised to tutor mathematics. Specifically, the topic of choice was "The Place Value of Digits in Whole Numbers." This topic is part of a bigger topic of Whole Numbers, and it is pivotal in equipping a learner with the foundations for reading, comparing, and performing arithmetic operations on whole numbers, making it crucial in understanding other more complex topics [19, 20]. We also chose this topic because of the minimal prior knowledge required to learn it, which is just being able to identify the digits 0–9. We placed considerable weight on the prerequisite mathematical concepts a pupil must have mastered to effectively learn a topic, as gaps in these prerequisites could introduce extraneous variables and influence the findings. [21].

2.3 Contextualising the Social Robot as a Tutor

The system architecture comprised of three subsystems working together to provide the Swahili-speaking social robot mathematics tutor. These subsystems, as can be seen in Fig. 1, are the server-side, the client-side, and the Aldebaran NAO robot.

2.3.1 The Server-Side

Computationally demanding processes ran server-side. The server dealt with the program logic behind the robot tutor by processing all the data and providing the final content. It also managed what was being presented on the client-side subsystem. Python 3.11.5 served as the core programming language to control the system logic. The system randomly generated two numbers at runtime, which were used as examples during the topic's introduction to the pupils. The first example was a three-digit number, demonstrating place values of "ones", "tens", and "hundreds". The second example had a number with four digits, introducing the maximum place value of "thousands". These were consecutively used to enable gradual learning.

Additionally, a function was created to generate exercises for pupils to try out after the robot introduced the topic. This aimed to (1) determine whether the learner had understood the topic, and (2) offer personalised feedback. The feedback included praise for correct answers and clarification with encouragement for incorrect answers, following advice from Koenka and Anderman [22], who emphasise the importance of personalised feedback in children's learning. For each question, the system drew a random number with four unique digits and selected one of the digits in the number to be used as a case for the pupil to identify its place value in the number. This ensured a virtually limitless variation of exercises.

For the introduction, narration of the questions, and provision of dynamic personalised feedback, we used OpenAI's GPT-3.5 API (*engine: gpt-3.5-turbo-instruct*) as our LLM of choice. GPT-3.5 has proven successful in similar settings and can interpret and generate Swahili content [15]. In contrast to classical Natural Language Processing (NLP) approaches that typically rely on predefined dialogue trees or handcrafted rules, the use of an LLM enabled flexible generation of context-aware explanations and personalised feedback based on pupils' responses. Furthermore, as LLMs were increasingly being explored in educational and conversational applications, we aimed to evaluate their feasibility for generating conversational mathematics content in a low-resource language such as Swahili.

In the prompts, we added contextual information to make the interactions more personalised and meaningful [23], and to control the theme and scope [24]. These contextual variables included the tutor's (NAO robot) name, *Asha* – which is a common female name in Tanzania; the tutor's title, "*mwalimu*", which is a Swahili word for tutor; the pupil's name; the pupil's age; the language for the interaction – Swahili; and the purpose of the interaction (in Swahili), which translates to – "*to learn about the place value*

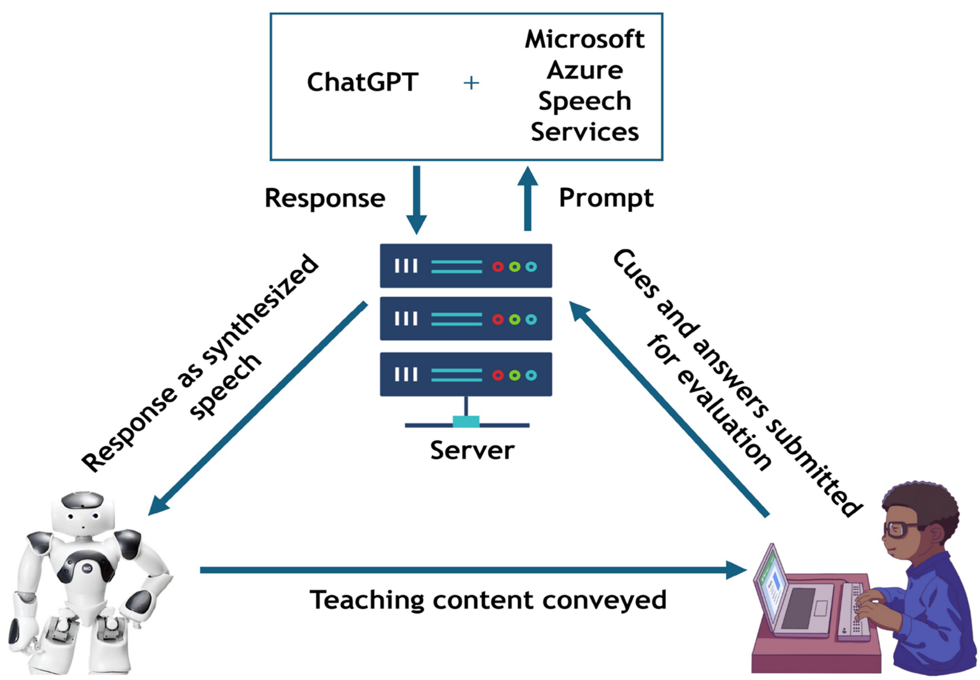


Fig. 1 System architecture

of digits in whole numbers"; and the locale – Tanzania. This ensures relevant and contextualised content.

However, after a few lab trials, we discovered that the output provided by GPT-3.5 sometimes used incorrect Swahili terminologies. For example, in the "ones" place value, it frequently used the word "*kitengo*", which translates as unit (in the context of a department/section), whereas the right Swahili word is "*mamoja*". This was also hinted at in a study by Panda and Kaur [25], that sometimes GPT-3.5 has a limited understanding of domain-specific jargon, which can result in incorrect responses. To remedy this, we created a custom English-to-Swahili dictionary (key-value pair) with contextualised translations of all the crucial terminologies, including the place value of digits and the right-left method of identifying them. While this limitation has been partially addressed in later GPT models, it remains significant in smaller models, particularly those deployed on local hardware, where additional measures are required to ensure context-appropriate terminology.

Finally, the dictionary and the contextual variables were both embedded in the prompts. The system utilised four types of prompts: (1) introduction prompt, the output of which was an introduction of the topic, stating the name, relevance, and real-world application of the topic, using the two generated examples for illustrations; (2) question prompt, which provided the textual narrative for the question to be attempted; (3) positive feedback prompt, whose output was a personalised praise when the pupil got a question right; and (4) encouragement prompt, whose output was a personalised encouragement and explanation when the pupil got the question wrong.

In addition, Microsoft Azure Speech services (Text-to-Speech) were used to convert the textual output generated by the LLM to a speech delivered by the robot. We selected a Tanzanian Swahili-speaking voice to be used when rendering the speech. This completed our transformation of the robot into an intelligent Swahili-speaking robot tutor.

Fig. 2 A pupil doing an exercise during the session

2.3.2 The Client-Side

The client-side subsystem served as the interactive interface between the pupils and the robot tutor, functioning as a digital blackboard where visual illustrations and exercises were displayed. It consisted of a touchscreen application running on a computer, enabling pupils to interact with the system in a tablet-like manner.

The application was developed using version 2.2.1 of the Kivy Python framework. Dynamic content, including examples and questions, was generated by the server-side subsystem and rendered on the application interface. The server also triggered the feedback displayed after a pupil attempted an exercise.

Following the robot's narrated introduction to the topic, pupils could proceed through the tutoring session by interacting with on-screen buttons labeled *Endelea* ("continue") and *Inatosha* ("I am done"). Selecting *Endelea* prompted the server to generate a new question asking the pupil to identify the place value of a digit in a number. The question was then displayed together with multiple-choice answers in Swahili corresponding to the place values.

Upon selecting an answer, pupils received immediate feedback from the NAO robot. Correct answers were acknowledged with praise, while incorrect responses triggered encouraging narration accompanied by a visual breakdown of the number into its respective place values to support understanding. Pupils could then continue attempting additional questions or end the session at any time (Fig. 2).

2.3.3 The NAO Robot

The NAO robot, programmed using Python 2.7.18 with built-in Naoqi modules, completed our setup by realising the physical embodiment of an intelligent AI-powered tutor. The NAO robot, acting as a tutor, rendered the speech that introduced the course, narrated the questions, and provided feedback, as relayed to it from the server-side. It also provided cues for the pupils to respond through the application's interface on the client-side.

2.4 The Experiment

The experiment was conducted at one school at a time from January to March, 2024. The targeted pupils were those in the third and fourth grades, often aged 9 and 10 years, respectively, as they are the ones who learn the topic with the ceiling of place values being thousands. In each school, six pupils were randomly selected to participate in the study, three coming from each grade.

2.4.1 Experimental Procedure

We started by having the pupils fill in their demographic information in a questionnaire. Then, they were individually administered a pre-test on the topic. In the pre-test, pupils were first required to write down digits 0–9 (as a basic screening task to ensure that they were familiar with numerals before proceeding to the place value assessment), followed by ten standard questions on identifying place–e.g., *what is the place value of 3 in 1324?*. The pre-test was meant to assess their knowledge before the intervention. And, because the topic had already been covered in class by their teachers, the pre-test score gave us the baseline for the learning attained from a human teacher, and a point of reflection for the improvements following being tutored by the NAO robot.

Thereafter, each pupil participated in an individual session with the NAO robot, where it introduced itself using its Swahili identity, introduced the topic, and guided them through a series of exercises and personalised feedback. Each pupil was required to attempt at least five exercises during the session, but they could do more as our system could generate limitless variations.

Following the session, the pupils completed a post-test with similar questions to the pre-test to assess their understanding after the intervention. Both the pre- and post-tests were graded by their teachers.

2.4.2 Qualitative Feedback

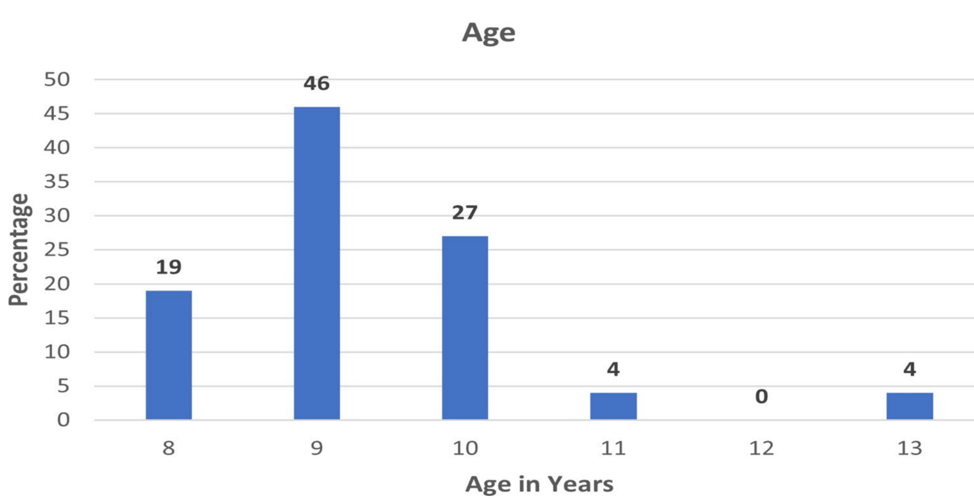
Subsequently, we conducted focus group discussions with the pupils (Fig. 3) to gather their views on the entire experience of having NAO as their tutor. In each school, the focus group discussion was conducted immediately after the post-test.

Fig. 3 Focus group discussion with the pupils

2.5 Study Participants

We recruited 30 pupils, however, results of four pupils could not be used due to technical difficulties during the experiments (described in 3.3), leaving 26 results for analysis. As presented in Fig. 4, the age of the pupils who participated in the study ranged between 8 and 13 years, the majority being 9 years old (46%), followed by 10-year-olds (27%). Contextually, the wide age range can be attributed to various socio-economic factors in Tanzania, leading to children not being enrolled in primary schools at the recommended starting age of 7. Additionally, some pupils have to repeat a grade due to unsatisfactory academic development [26, 27]. Fifty-four percent (54%) of the pupils were in the fourth

Fig. 4 Pupil's age



grade, and the rest were in the third grade. Also, 58% of the pupils were females, while 42% were males.

2.6 Analysis

For the test scores, a Shapiro-Wilk test was performed and showed that the distribution of the scores in the pre- and post-tests departed significantly from a normal distribution ($p<0.05$). Therefore, we used the Wilcoxon signed-rank test to determine whether there was a significant difference between the pupils' results in the pre- and post-tests. The data from the focus group discussions were thematically analysed to identify the collective themes from pupils' views. The pupils' demographic data were analysed descriptively.

3 Results

3.1 NAO Tutoring the Mathematics Topic

3.1.1 Excerpts from NAO's Tutoring

The NAO robot, powered by GPT-3.5, autonomously tutored the pupils without the need for intervention by an operator. It began by introducing the topic, followed by exercises and personalised feedback. In this section, we provide two examples of each, translated from the original Swahili. The names Neema (female) and Baraka (male) have been used to ensure anonymity while preserving the original meaning. These excerpts illustrate the type of interaction delivered during the intervention.

Introduction 1: *"Hi Baraka! Welcome to our math class. Today we are learning about the place value of digits in whole numbers. This is very important as it will help you understand how numbers work and how to organise them. As you can see in your table, in the number 607, digit 7 is in the ones place, digit 0 is in the tens place, and digit 6 is in the hundreds place. Likewise, in the number 2,179, digit 9 is in the ones place, digit 7 is in the tens place, digit 1 is in the hundreds place, and digit 2 is in the thousands place. Remember to always use the counting method starting from the right-hand side to the left to be able to know the place value of the digits in the number. For more examples, click the word 'Continue'. I hope you enjoy our class and learn a lot!"*

Introduction 2: "*Hello Neema, my name is Ms. Asha and I am your tutor. Today we will learn about the place value of digits in whole numbers. This is very important because it will help you understand whole numbers and count accurately. As you can see on your screen, in the first example, we have the number 830. In this number, 0 is in the ones place, 3 is in the tens place, and 8 is in the hundreds place. For the second example, we have the number 4,793. Here, 3 is in the ones place, 9 is in the tens place, 7 is in the hundreds place, and 4 is in the thousands place. Remember to always use the counting method from right to left. This will help you place each digit in the correct position. So now, click 'Continue' for more examples and remember to use the table to list the value of the digits in whole numbers.*"

Question 1: "*Remember to carefully consider the place value of the digit 8 in the number 7,738. You can choose between "Ones," "Tens," "Hundreds," or "Thousands." Please use the counting method from right to left. Baraka, you can succeed, please choose the correct answer.*"

Question 2: "*Find the place value of the digit 5 in the number 5,723. Choose the correct answer: "Hundreds," "Thousands," "Tens," or "Ones." Remember to count from right to left. Continue with more examples in the table.*"

Positive feedback 1: "*Hello Neema! I'm glad you were able to find the correct answer to my question. Let's always remember to use the place value of digits to count whole numbers. You can click 'Continue' or 'Done' if you have finished learning for today. Thank you!*"

Positive feedback 2: "*God bless you, Baraka! You chose 'ones', which is the place value of 8 in the number 7738. You have understood the topic well. Click 'Continue' for more questions or 'Done' if you're finished for today. Welcome again!*"

Negative feedback 1: "*I'm sorry, Baraka! The place value of 3 in the number 6,683 is 'ones.' You should use the counting method from right to left. You may also refer to the table on the screen. You can click 'Continue' for more questions or 'Done' if you're done. Welcome to learn more about the place value of digits in whole numbers, Baraka.*"

Negative feedback 2: "*Neema, you were to find the place value of 5, in the number 4,857. However, you answered 'thousands'. The value of 5 is in the 'tens' place. You can see this by comparing it to the table on your screen which shows the correct value of the digit in the number, counting from right to left. Do not use any other method.*"

Moreover, from our observation, there was at times mispronunciation of words, specifically, the Swahili word for 'Value', which is '*Thamani*'. The th- should be pronounced as in THing (phonetically θ), but NAO pronounced it dh- as in THAt (phonetically $/ð/$). Also, there were occassional instances (unquantified) when the system produced semantically incorrect and/or incomprehensible output.

3.1.2 Quantitative Results

The study yielded a positive result for the robot tutoring. The NAO robot's tutoring, exercises, and personalised feedback resulted in an increased learning outcome. As shown in Fig. 5, the mean scores significantly improved

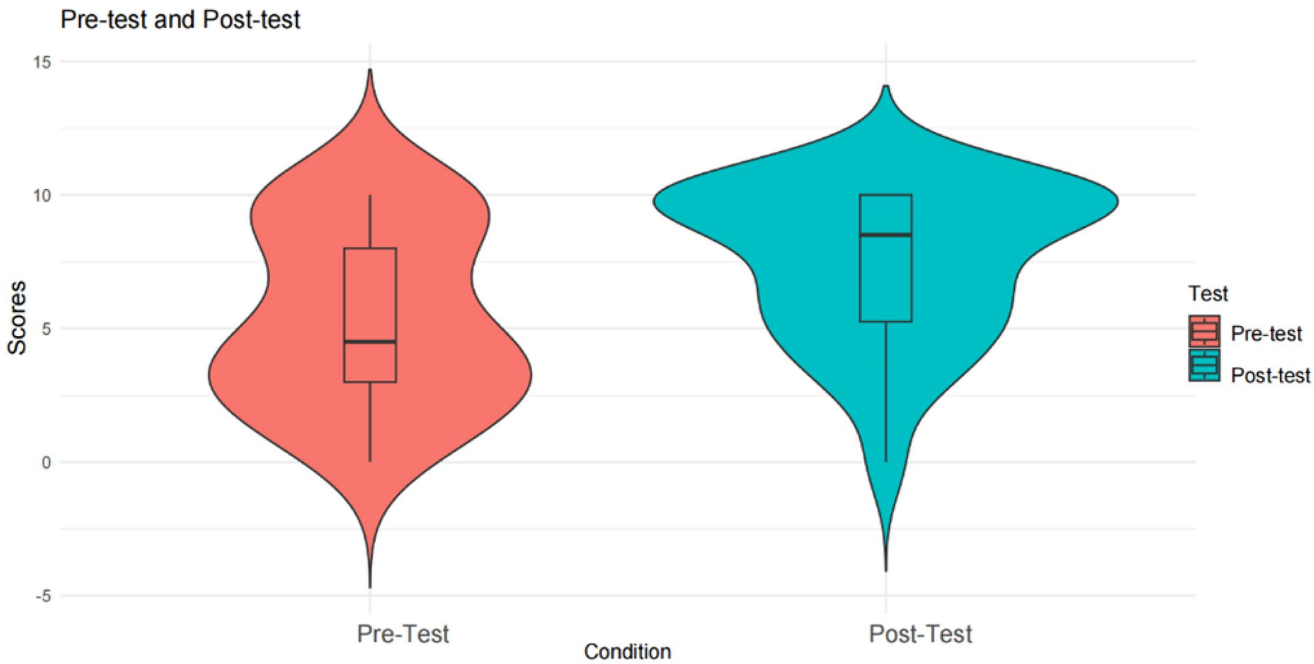


Fig. 5 Pre- and post-test scores

from 5.39 (SD=3.3) to 7.52 (SD=2.9) out of 10 for the pre- and post-tests, respectively. The Wilcoxon signed-rank test with continuity correction confirms this significance (Z=3.628, p-value<0.001). Following the recommendations for non-parametric data [28], the effect size was calculated as $r = \frac{Z}{\sqrt{N}}$, where N is the number of non-zero paired differences used to compute the Z statistic. A large effect size ($r = 0.88$) was observed, indicating a robust impact of robot tutoring on pupils' performance.

Note that the pupils had received a lesson on the same topic in class prior to being tutored by the robot. As such, the results reported here reflect the improvement observed after the additional tutoring session with the robot.

3.2 Pupils' Feedback from the Experience

3.2.1 Reflection on their Interaction with NAO as their Tutor

Prior to the experiment, we assessed the pupils' knowledge of the concepts of AI and social robots. Only 8% of the pupils had heard of AI, and 15% were familiar with social robots, mainly through media such as movies, television, and cartoons. For all pupils, this was their first real-life experience interacting with a social robot.

In the focus group discussion conducted after the intervention, pupils expressed overwhelmingly positive feelings towards the NAO robot. Their comments primarily highlighted its friendliness, tutoring ability, and clear explanations. Common responses included: "*It is friendly*", "*it is knowledgeable*", "*it knows how to teach*", "*it has a nice voice*", "*it gives a lot of examples*", and "*it is beautiful – nice shape and colour*".

Regarding the perceived learning support, pupils affirmed that the NAO robot helped them understand the topic better. They commended its consistent emphasis on using the right-hand side to left-hand side counting method when determining place values. Some of their quotes were:

"*It helped me understand the topic because it explained in a friendly way and provided many examples.*"

"*Yes, I understood because it constantly insisted that I count the place values from the right-hand side to the left.*"

When discussing NAO's role in the learning process, all pupils preferred its style to be that of a tutor rather than a peer or novice. The majority (85%) favoured a friendly tutor personality. However, for the few who suggested it to be strict, interesting reasons were provided, such as:

"*If it is too friendly some pupils will not do their exercises and homework.*"

"*Some pupils might play with it instead of learning.*"

A few pupils even proposed that the robot should be able to punish the pupils in case of laziness and misconduct. This can be understood in the context of current pedagogical practice in Tanzania, where classroom management is strict and assertive, with corporal punishment still being prevalent in primary schools [29].

3.2.2 Other Topics and Subjects that the NAO Robot Should Tutor

The pupils were asked to mention other mathematics topics or skills they would like the NAO robot to tutor them. Arithmetic and solving word problems were proposed by the majority. Also, when asked to propose other subjects they would like the NAO robot to tutor them, English was the most proposed subject, followed by Swahili and Social Studies.

3.3 Implementation Challenges

Although to a large extent the experiment was conducted successfully, several situational and infrastructural challenges were observed during implementation in the schools. Firstly, none of the schools had internet infrastructure installed; therefore, we had to rely on mobile hotspots. Nevertheless, while connectivity was sufficient in urban schools, it was unstable in rural schools. Secondly, conversational latency–further affected by unstable connections–made the interaction less natural. Lastly, in one school we experienced electricity outages that directly affected the implementation. As a result of these challenges, four pupils were excluded because they were unable to complete their sessions successfully. These challenges and practical considerations are discussed in detail in [30].

4 Discussion

4.1 Awareness of AI and Social Robots Among Pupils in Public Primary Schools

The results highlight a general lack of awareness regarding AI and social robots among pupils in Tanzanian public primary schools. This suggests limited exposure to these technologies in their educational or home environments. This aligns with existing literature on the technology gap in Tanzania and the insufficient utilisation of technology in education [31]. Moreover, most of the studies and interventions related to social robots and human-robot interaction are based in Europe, Asia, and North America, as compared to Africa [14, 32, 33]. In addition, most of the AI-based research in Africa is geared towards Agriculture, Healthcare, and FinTech [34]. This underscores a significant opportunity for research and the use of AI and social robots to improve the education sector in Tanzania.

4.2 NAO Robot as a Swahili-Speaking Mathematics Tutor

4.2.1 The Use of a Custom Dictionary and Contextual Information

Swahili is a low-resource language, with fewer datasets available for training LLMs as compared to more widely used languages. Moreover, GPT-3.5 still exhibits notable limitations when processing Swahili [15]. This surfaced vividly in our study, as despite our efforts, content generation still occasionally resulted in incorrect or nonsensical outputs.

However, in the case of the wrong use of terminologies, our approach of embedding a custom English-Swahili dictionary containing context-specific translations for crucial terms related to the topic proved effective in ensuring accurate terminology use in the generated content. Refining prompts to better fit the context and to yield an output more fitting and specific to one's intentions has also proven effective in similar studies [35, 36].

In a similar vein, including contextual information such as the name and age of the pupil made the interaction more personalised and age-appropriate, as can be seen in the examples provided in 3.1.1. This affirms the findings in a study by Baxter et al. [23]. Moreover, providing the NAO robot's name also made the introduction more friendly and natural as practised by human teachers, where they introduce themselves on the first encounter to create rapport. Interestingly, adding the country information, Tanzania, led the LLM to adopt a more culturally relevant tone, even using the Swahili phrase "*Mungu akubariki*" (God bless you) (see positive feedback 2 in 3.1.1). Phrases related to God are deeply woven into Tanzanian culture and are commonly used in various scenarios and circumstances that are not necessarily related to religion [37–39]. The cultural relevance of the contextual information to refine prompts reinforces their effectiveness.

4.2.2 Personalised Feedback and the Robot as a Remedial Tutor

The findings also reveal that the NAO robot provided personalised feedback to the learners. It consistently addressed pupils by their names, and provided tailored corrections and encouragement in response to incorrect answers. In view of future applications of educational social robotics, this may help address challenges associated with providing personalised feedback in large classes during normal classroom sessions. The robot may therefore complement classroom instruction, particularly for learners who require additional support. This interpretation is supported by the observed increase in mean post-test scores, along with a reduction in the standard deviation, suggesting not only improved performance but also more consistent understanding across pupils.

The findings also highlighted several situational and infrastructural challenges that need to be rectified for smooth deployment of robots in schools. However, these aspects have been discussed in detail in [30] and are therefore not elaborated further here.

4.2.3 Friendly Vs Strict NAO's Personalities

Our study utilised a friendly and engaging NAO personality, which included the use of the pupil's name, applause, and encouragement. Knowing that it was talking to a child who wants to learn also made the robot supportive and considerate. While this yielded positive results in the experiment, we obtained cautious remarks from the pupils in the qualitative part of the study. They alerted us that the over-friendly nature of the NAO robot (the tutor) could negatively affect learning as it might impair the seriousness and concentration of some learners. Their suggestions are in line with the findings by Kennedy, Baxter, and Belpaeme [40], where an asocial robot was more effective than a social one. Thus, for maximum impact, it might be crucial to balance these personalities and have them adaptive according to the learner.

4.3 Cognitive Learning Gains and Tutoring Methods

The study indicates that the use of the NAO robot as a mathematics tutor following classroom instruction was associated with improved pupil understanding of the topic. The observed effect size ($r=0.88$) indicates a substantial practical improvement in performance, suggesting that the intervention had a meaningful educational impact beyond statistical significance. These findings are consistent with previous studies conducted in other contexts [14, 40, 41]. From the pupils' perspectives, the elements that attributed to the observed improvement were the NAO robot's use of sufficient examples, tailored feedback when they got a question wrong, calm voice, appealing physical appearance, and overall considerate personality. Thus, these attributes, when embodied cautiously, might help other similar interventions to achieve a positive cognitive learning impact.

5 Conclusion, Limitations, and Future Work

The study set out to explore the use of an LLM-powered Swahili-speaking NAO robot as a Mathematics tutor following classroom lesson by a human teacher in Tanzanian Primary schools. The findings provide answers to the three research questions.

i. **Adapting LLMs for Swahili primary school math tutoring in Tanzania:** The study demonstrates that LLMs (and specifically GPT-3.5 and later versions) can be leveraged to create educational content for Mathematics in Tanzania in the Swahili language, despite it being a low-resource language for LLM training. However, the model still needs assistance to ensure the use of correct terminologies and jargon specific to the topic. Thus, incorporating a custom dictionary and contextual information is crucial.
ii. **Improvements in pupils' understanding of the math topic:** The intervention was associated with improved pupils' understanding of place values, as reflected in the significant increase in pre- and post-test scores. These findings indicate that a robot tutor can support learning after classroom instruction. Furthemore, the results highlight the potential of robots acting as remedial tutors, offering one-on-one personalised tutoring.
iii. **Pupils' perception of NAO as a tutor:** Pupils unanimously expressed positive feedback regarding NAO's role as a tutor. They appreciated its friendly demeanor, clear explanations, and personalised feedback. Additionally, they expressed interest in having NAO as their tutor in other topics and subjects.

While the pilot study was successful and the focus group discussions illuminating, we acknowledge that the views and performances of 26 pupils from 5 schools do not suffice to draw conclusions for a country with more than 19,000 schools [17] and great demographic and cultural diversity. As this was a proof-of-concept study, these findings should be interpreted as exploratory evidence from a limited deployment rather than as broadly generalisable results. For generalisability, we recommend a broader study or one that tests similar parameters in a different demographic setting in the country and refer the reader to a follow-up study [42].

Moreover, studies should be conducted with a focus on other topics and subjects, some of which were proposed by the pupils. Exploring different robot behaviours/personas could also yield interesting results. As for adapting LLMs (such as GPT-3.5, later versions, and especially edge-run LLMs) to better suit the Tanzanian context and the Swahili language, further experiments should be conducted to detect and analyse errors in the generated output and to enrich the training data.

Additionally, the significant cognitive and affective benefits our study demonstrated were obtained in brief one-time sessions. Research should be conducted to examine if this impact is sustained over prolonged periods of exercises and for tasks with varying complexity. This is stressed in a recent cross-cultural and socio-economic study, where pupils from comparatively lower socio-economic backgrounds displayed more positive attitudes during a robot intervention, but the effects diminished more quickly afterwards [43]. The study highlights the need for longer-term investigations, as the improvements observed in our study may partly be influenced by a novelty effect associated with the introduction of a social robot in Tanzanian school settings.

Finally, Seaborn et al. [44] address the over-reliance on WEIRD (Western, Educated, Industrialised, Rich, and Democratic) populations in Human-Robot Interaction (HRI) research, which has limited the generalisability of findings across diverse cultural contexts. Our research is one of few using HRI in a non-WEIRD context, helping to create a more comprehensive understanding of how different cultural backgrounds potentially influence interactions with robots.

Acknowledgements Special thanks to VLIR-UOS for funding this study through the Mzumbe University IUC programme and the Ghent University Global Minds scheme. Our heartfelt gratitude also extends to all the participants whose dedicated time and invaluable input were essential to the success of this research. Finally, we thank the government of Tanzania for granting us all the necessary permits, which made the process smooth.

Data Availability The dataset generated and analysed in this study is not publicly available due to the privacy of the participants and the terms of the research permit. However, depending on the nature of the request, it may be made available by the corresponding author.

Declarations

Ethics Approval The study was endorsed by Mzumbe University (reference: MU/PF/2650/54), which supported the application for research clearance by introducing the study objectives, design, and researchers to the relevant authorities. Formal approval was subsequently granted by the Tanzanian government through the President's Office – Regional Administration and Local Government (PO-RALG), which oversees public primary schools (permit number AB.307/323/01/40), allowing the study to be conducted in schools. Participants were informed about the study and their right to withdraw at any time without any negative consequences. Furthermore, as the study involved pupils, their teachers were present throughout the sessions, providing non-intrusive supervision. Permission to publish the results of the study was also obtained (permit number AB.307/323/01/228).

Conflict of Interest The authors declare no conflict of interest.

References

1. Olasoji OV, Felicity UU, Onoh DO (2023) The role of mathematics education in achieving sustainable development goals (SDGs). ESUT J Educ (EJE) 6:247–255
2. Sun X, Yang Y (2022) Advanced mathematics in vocational universities. JCER 6(6):13–18. https://doi.org/10.26689/jcer.v6i6.4006
3. Mazana MY, Montero CS, Casmir RO (2020) Assessing students' performance in mathematics in Tanzania: the teacher's perspective. INT ELECT J Math Ed 15(3):0589. https://doi.org/10.29333/iejme/7994
4. Kabote SJ, Niboye EP, Nombo CI (2014) Performance in mathematics and science subjects: a gender perspective for selected primary schools in rural and urban Tanzania. IJGWS 2(3):87–105. https://doi.org/10.15640/ijgws.v2n3a6
5. Bakar AS, Mwila PM (2022) Teacher–student ratio: the status quo in public secondary schools in Ubungo Municipality, Tanzania. AJESS 24–36. https://doi.org/10.9734/ajess/2022/v34i4741
6. Ndijuye LG (2023) Early learning attainments of children of naturalized citizens of refugee backgrounds in the sub-saharan region: evidence from Tanzania. Child Dev Perspect 17(1):67–73. https://doi.org/10.1111/cdep.12479
7. Baraza la Mitihani la Tanzania: Taarifa ya Matokeo ya Mtihani wa Kumaliza Elimu ya Msingi (PSLE) (2023) Uliofanyika Septemba. https://www.moe.go.tz/sites/default/files/PRESS%20PSLE%202023%20final_231123_121622_1.pdf
8. Casheekar A, Lahiri A, Rath K, Prabhakar KS, Srinivasan K (2024) A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: applications, open challenges and future research directions. Comput Sci Rev 52. https://doi.org/10.1016/j.cosrev.2024.100632
9. Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Science 381(6654):187–192. https://doi.org/10.1126/science.adh2586
10. Lubowitz JH (2023) ChatGPT, an artificial intelligence chatbot, is impacting medical literature. Arthrosc - J Arthroscopic Relat Surg 39:1121–1122. https://doi.org/10.1016/j.arthro.2023.01.015
11. Verhelst E, Demeester T, Janssens R, Belpaeme T (2024) Adaptive second language tutoring using generative AI and a social robot. In Companion of the 2024 ACM/IEEE International Conference on HumanRobot Interaction (HRI'24 Companion),,, IEEE Computer Society, Boulder, CO, USA, 1080–1084. https://doi.org/10.1145/3610978.3640559
12. Jauhiainen JS, Guerra AG (2023) Generative AI and ChatGPT in school children's education: evidence from a school lesson. Sustainability (Switz) 15(18):14025. https://doi.org/10.3390/su151814025
13. Pesovski I, Santos R, Henriques R, Trajkovik V (2024) Generative AI for customizable learning experiences. Sustainability (Switz) 16(7):3034. https://doi.org/10.3390/su16073034
14. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. Sci Robot 3(21):5954. https://doi.org/10.1126/scirobotics.aat5954
15. Liao C, Wu X (2023) On the evaluation of chatgpt-3.5 on swahili classification tasks. Institute of Electrical and Electronics Engineers Inc, pp 25–28. https://doi.org/10.1109/AIIIP61647.2023.00011

16. Common crawl foundation: statistics of common crawl monthly archives: distribution of languages (2024) Accessed: 2024-05-20. https://commoncrawl.github.io/cc-crawl-statistics/plots/languages
17. National Bureau of Statistics: Tanzania in figures (2022). https://www.nbs.go.tz/index.php/en/tanzania-in-figures/891-tanzania-in-figures-2022
18. Mtewele GC (2012) The influence of traditional customs and practices on girls' secondary education in Morogoro region in Tanzania a case study. Master's thesis, University of Oslo
19. Schmittau J, Vagliardo JJ (2006) Using concept mapping in the development of the concept of positional system 590–597. https://citeseerx.ist.psu.edu/document?repid=rep1%26type=pdf%26doi=e2ab7652df6e5ff6ed6063458e94f468708b4376
20. Nataraj MS, Thomas MOJ (2007) Developing the concept of place value 523–532. https://citeseerx.ist.psu.edu/document?repid=rep1%26type=pdf%26doi=0e860f7c1bbf0f3c8bf13e80dc7b32e79a3345f8#page=526
21. Lynch-Arroyo R, Asing-Cashman J (2016) Using edutainment to facilitate mathematical thinking and learning: an exploratory study. Pac J Math Educ Educ All 9:37–52
22. Koenka AC, Anderman EM (2019) Personalized feedback as a strategy for improving motivation and performance among middle school students. Middle Sch J 50(5):15–22. https://doi.org/10.1080/00940771.2019.1674768
23. Baxter P, Ashurst E, Read R, Kennedy J, Belpaeme T (2017) Robot education peers in a situated primary school study: personalisation promotes child learning. PLoS One 12(5), 0178126. https://doi.org/10.1371/journal.pone.0178126
24. Henrickson L, Meroño-Peñuela A (2025) Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI Soc 40(2):903–918. https://doi.org/10.1007/s00146-023-01752-8
25. Panda S, Kaur N (2023) Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers. LHTN 40(3):22–25. https://doi.org/10.1108/LHTN-02-2023-0032
26. Kobayashi C (2016) Determinants of primary school attendance in rural Tanzania: an analysis of children in and out of school. Africa Educ Res J 7:137–154
27. Benjamin GR (2007) Implementation of the universal primary education policy and its effectiveness in Tanzania. Master's thesis, Kampala International University. https://ir.kiu.ac.ug/items/4f0bbbe0-31a1-4bf9-9180-a715000cf680,
28. Field A, Miles J, Field Z (2012) Discovering statistics using R. SAGE Publications, London
29. Kalolo JF, Kapinga OS (2023) Towards ending corporal punishment in African countries: experiences from Tanzania. Int J Educ Dev 102:102839. https://doi.org/10.1016/j.ijedudev.2023.102839
30. Rutatola EP, Stroeken K, Belpaeme T (2025) "Habari, colleague!": a qualitative exploration of the perceptions of primary school mathematics teachers in Tanzania regarding the use of social robots. Appl Sci 15(15):8483. https://doi.org/10.3390/app15158483
31. Tandika PB, Ndijuye LG (2019) Pre-primary teachers' preparedness in integrating information and communication technology in teaching and learning in Tanzania. ILS 121(1/2):79–94. https://doi.org/10.1108/ILS-01-2019-0009
32. Lim V, Rooksby M, Cross ES (2021) Social robots on a global stage: establishing a role for culture during human–robot interaction. Springer. https://doi.org/10.1007/s12369-020-00710-4
33. Papakostas GA, Sidiropoulos GK, Papadopoulou CI, Vrochidou E, Kaburlasos VG, Papadopoulou MT, Holeva V, Nikopoulou VA, Dalivigkas N (2021) Social robots in special education: a systematic review. Electronics 10(12):1398. https://doi.org/10.3390/electronics10121398
34. Sampene AK, Agyeman FO, Brenya R, Wiredu J (2022) Artificial intelligence as a path way to Africa's transformations. J Educ Chang Multidiscip Eng Sci Technol (JMEST) 9:14939–14951
35. Lee U, Jung H, Jeon Y, Sohn Y, Hwang W, Moon J, Kim H (2024) Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. Educ Inf Technol 29(9):11483–11515. https://doi.org/10.1007/s10639-023-12249-8
36. Wang L, Bi W, Zhao S, Ma Y, Lv L, Meng C, Fu J, Lv H (2024) Investigating the impact of prompt engineering on the performance of large language models for standardizing obstetric diagnosis text: comparative study. JMIR Form Res 8:e53216. https://doi.org/10.2196/53216
37. Sheridan D (2018) "If you greet them, they ignore you": Chinese migrants, (refused) greetings, and the interpersonal ethics of global inequality in Tanzania. Anthropological Q 91(1):237–265. https://doi.org/10.1353/anq.2018.0007
38. Namwata P (2012) Categories of street beggars and factors influencing street begging in central Tanzania. Afr Study Monogr 33:133–143. https://doi.org/10.14989/158998
39. Lindhardt M (2010) 'If you are saved you cannot forget your parents': agency, power, and social repositioning in Tanzanian born-again Christianity. J Religion Afr 40(3):240–272. https://doi.org/10.1163/157006610X530330
40. Kennedy J, Baxter P, Belpaeme T (2015) The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning, vol. 2015-March. IEEE Computer Society, Portland, Oregon, USA, pp 67–74. https://doi.org/10.1145/2696454.2696457
41. Lee H, Lee JH (2022) The effects of robot-assisted language learning: a meta-analysis. Educ Res Rev 35:100425. https://doi.org/10.1016/j.edurev.2021.100425
42. Rutatola EP, Ntahomvukye EC, Stroeken K, Belpaeme T (2026) Robot tutors or peers? Evaluating math learning and conformity with LLM-powered robots in Tanzanian primary schools. In Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction, pp 237–245
43. Deshmukh A, Cross ES, Foster ME (2025) Leveraging social robots to promote hand hygiene: a cross-cultural and socio-economic study of children in diverse school settings. In: HRI'25: proceedings of the 2025 ACM/IEEE international conference on human-robot interaction (March 2025). ACM/IEEE. https://doi.org/10.5555/3721488.3721536
44. Seaborn K, Barbareschi G, Chandra S (2023) Not only weird but "uncanny"? a systematic review of diversity in human–robot interaction research. Int J Soc Robot 15(11):1841–1870. https://doi.org/10.1007/s12369-023-00968-4

Edger P. Rutatola is a PhD researcher in the Artificial Intelligence and Robotics (AIRO) research group at Ghent University–imec, Belgium. He is also an Assistant Lecturer and researcher at the Centre for AI Research and Innovation (CAIRI) at Mzumbe University, Tanzania. His work focuses on contextualising AI innovations to address societal challenges in low-resource settings. In particular, his current research explores the use of AI and robotics to enhance teaching and learning in Tanzanian primary schools, addressing technical, infrastructural, and societal constraints to ensure these innovations are suitable and effective.

Qiaoqiao Ren is a Postdoctoral Researcher with the Interaction Design Group, KTH Royal Institute of Technology, Stockholm, Sweden. She received the Ph.D. degree from IDLab-AIRO, Ghent University–imec, Ghent, Belgium. Her research focuses on affective touch and human–robot interaction, with an emphasis on multimodal perception, tactile sensing, and emotion communication through touch and sound. Her work explores how robots can both perceive and express emotions through embodied interaction.

Maria Jose Pinto-Bernal is a Biomedical Engineer and PhD researcher at Ghent University – imec, Belgium. Her research focuses on human–robot interaction, large language model integration in social robots, and multimodal conversational systems designed to support long-term social interaction and emotional wellbeing in healthcare contexts.

Koen Stroeken is a full professor of Africanist anthropology at Ghent University and co-director of the research centre CARAM. He has published several papers on African cultural dynamics, including the peerreviewed books Moral Power (Berghahn), Medicinal Rule (Berghahn), and Simplex Society (Springer).

Tony Belpaeme is a Professor at Ghent University (Belgium) and a Senior Researcher at imec. He leads a team studying cognitive robotics and human-robot interaction. Starting from the premise that intelligence is rooted in social interaction, the team seeks to further the science and technology behind artificial intelligence and social robotics. This results in a spectrum of results, from theoretical insights to practical applications.