SUKHSANDESH: An Avatar Therapeutic Question Answering Platform for Sexual Education in Rural India

Salam Michael Singh[1] , **Shubhmoy Kumar Garg**[3] , **Amitesh Misra**[3]
Aaditeshwar Seth[3,4,5] and **Tanmoy Chakraborty**[1,2]

[1]Department of Electrical Engineering, Indian Institute of Technology Delhi, India
[2]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, India
[3]Gram Vaani Community Media (Onion Dev Technologies Pvt. Ltd.), New Delhi, India
[4]Department of Computer Science and Engineering, Indian Institute of Technology Delhi, India
[5]School of Information Technology, Indian Institute of Technology Delhi, India
mike95@iitd.com, {shubhmoy.kumar, amitesh.mishra}@oniondev.com, aseth@cse.iitd.ac.in,
tanchak@iitd.ac.in

Abstract

Sexual education aims to foster a healthy lifestyle in terms of emotional, mental and social well-being. In countries like India, where adolescents form the largest demographic group, they face significant vulnerabilities concerning sexual health. Unfortunately, sexual education is often stigmatized, creating barriers to providing essential counseling and information to this at-risk population. Consequently, issues such as early pregnancy, unsafe abortions, sexually transmitted infections, and sexual violence become prevalent. Our current proposal aims to provide a safe and trustworthy platform for sexual education to the vulnerable rural Indian population, thereby fostering the healthy and overall growth of the nation. In this regard, we strive towards designing SUKHSANDESH, a multi-staged AI-based Question Answering platform for sexual education tailored to rural India, adhering to safety guardrails and regional language support. By utilizing information retrieval techniques and large language models, SUKHSANDESH will deliver effective responses to user queries. We also propose to anonymise the dataset to mitigate safety measures and set AI guardrails against any harmful or unwanted response generation. Moreover, an innovative feature of our proposal involves integrating "avatar therapy" with SUKHSANDESH. This feature will convert AI-generated responses into real-time audio delivered by an animated avatar speaking regional Indian languages. This approach aims to foster empathy and connection, which is particularly beneficial for individuals with limited literacy skills. Partnering with Gram Vaani, an industry leader, we will deploy SUKHSANDESH to address sexual education needs in rural India.

1 Introduction

Sexual education strives toward building a person's lifelong sexual health by teaching them attitudes, values, and knowledge about intimacy, relationships, and identity [Ismail et al., 2015]. Sexual health is regarded as a state of physical, emotional, mental, and social well-being rather than merely the absence of illness or infirmity as reported by the WHO[1]. Moreover, sociocultural and psychological factors can increase the likelihood of effectiveness in delivering this education. Sexual education is crucial for adolescents as this is the phase in which they undergo hormonal changes[2]. However, sexual education is often stigmatised and thus becomes a hurdle in counselling vulnerable people. Furthermore, healthcare professionals often overlook this issue. For instance, they do not collect a proper history of sexual activities and do not openly discuss the issues due to cultural and traditional norms in society. Notably, there are cases where parents, politicians and teachers have opposed sexual education in Indian states, which include Chhattisgarh, Madhya Pradesh, Rajasthan, Karnataka, Gujarat and Maharashtra [Ismail et al., 2015]. They argued that sexual education leads to promiscuity and devalues the Indian ethos and cultural traditions. This will require a tremendous amount of ground-level awareness as such beliefs are deeply rooted in the society irrespective of their background. Moreover, young people are confused by a combination of myths, stigma, ignorance, social disparity, and misinformation. These lead to very critical issues such as sexually transmitted infection (STI), HIV/AIDS, early pregnancy, unsafe abortions, and sexual violence at an alarming rate[3].

1.1 Sexual Education: An Indian Perspective

Sexual education and sex-related topics are still taboo in India. A study among school students from the Ambala district in Haryana in India [Kumar et al., 2017] revealed that

---

[1]https://www.who.int/publications/i/item/978924151288
[2]https://www.who.int/publications/i/item/9789241501552
[3]https://www.who.int/publications-detail-redirect/WHO-RHR-10.12

93.5% favour sex education. 86.3% said sex education could prevent the occurrence of AIDS. 91.5% of adolescents preferred doctors to give them sex education, followed by teachers. However, many were hesitant to talk with their parents. Another work in the rural region of Bengaluru in India conducted a study among 981 students across 6th to 10th grades [Mehrolia *et al.*, 2023]. The study reports that the sexual education program positively impacted the students' confidence. Further, this fostered them towards hygiene and decision-making skills. Another study [Ray *et al.*, 2012] on the Bengali-speaking Hindu ethnic group from rural and urban demographics. The study was carried out among 101 rural 119 urban adolescent boys. They found that the demographics played a key role in the sexual awareness. The urban boys were more aware, and they were less likely to be involved in unprotected and penetrative sexual activities. Furthermore, the study revealed the role of media and schools in promoting healthy sexual behaviours among adolescents. Another work [Ghule and Donta, 2008], also found that rural people are more vulnerable towards unprotected sexual activities.

1.2 Sexual Education is Crucial

For every country, a sexually educated and well-aware population will lead to several benefits such as health, self-confidence and overall development of the economy [David *et al.*, 2013]. At the global level, adolescents, the most vulnerable group, account for 18% of the world's population as of 2009 [Khubchandani *et al.*, 2014]. And India records the largest number of adolescents, with 243 million counts. Because India is a developing country, timely sexual education is crucial for the overall development of the country [Tripathi and Sekher, 2013]. If left unattended, this will lead to various risks such as sexually transmitted diseases, premature pregnancies, unsafe abortions, sexual abuse and violence, and overall influence on health and well-being. Hence, there is an urgent need to spread awareness and foster a healthy future for the Indian population [Datta and Majumder, 2012]. Additionally, there are societal stigmas and taboos around sex-related topics in India. Therefore, a ground-level social awareness program is the need of the hour to reduce the stigmas prevailing in sexual education; then, the actual education will be accessible.

1.3 Existing Studies and Research Gaps

This work proposes a multi-staged avatar therapeutic Question Answering (QA) system for sexual education in rural India. Currently, a single system with both avatar therapy and QA for sexual education is not explored. In this section, the related work on QA for sexual education is discussed first, followed by studies on avatar therapy.

SnehAI [Wang *et al.*, 2022] is a chatbot in Hinglish codemixed data. This platform provides a safe and trustworthy environment for users. The authors targeted adolescent Indian girls to advocate for sexual and reproductive health. A WhatsApp-based bot named Wulu [Agarwal *et al.*, 2021] is reported, which focuses on the gender bias issue in Delhi. A work using a chatbot as a sexual health counsellor is reported [Mills *et al.*, 2023]. The chatbot is centred around sexual and reproductive health counselling by facilitating a safe and private platform to build a sense of trustworthiness for vulnerable users. Nadarzynski *et al.* [Nadarzynski *et al.*, 2021] conducted a chatbot-based study on women from the UK. They found that most users preferred this chatbot for anonymous education purposes and less suitable where empathy is required. Another chatbot [Bonnevie *et al.*, 2021] focused on contraception for young women from Black and Hispanic backgrounds. The study was conducted with 31 women in the age group of 16 to 25 years. The chatbot received a total of 4390 messages. This chatbot shared vital information to boost young women's confidence in contraception. Similarly, a chatbot for fertility awareness is reported [Maeda *et al.*, 2020]. Their awareness program showed significant gains in fertility knowledge.

Avatar-based AI has widely been used in the domains of mental health and cognitive behavioural therapy. An avatar-based chatbot [Goonesekera and Donkin, 2022] is reported for mental health anxiety during the COVID-19 pandemic. Their avatar is not a real-time text-to-audio system but an icon to soothe the users. On the other hand, the use of an avatar for combating auditory is used as an avatar therapy for schizophrenia [Holohan and Fiske, 2021]. As of now, the avatar has been predominantly used in the mental health domain for therapy purposes, specifically for auditory hallucinations [Holohan and Fiske, 2021].

Hitherto, as discussed above, these chatbots build upon the idea of providing safety and privacy. However, they lack certain aspects, such as these models not being robust to new queries and the responses being fixed as per the training data. Moreover, they lack the scalability for incremental training based on new training instances, and their safety measures are based on user identity anonymisation only without additional guardrails. Furthermore, the avatar is limited to therapy purposes only, while this has several other potential in terms of an empathetic educational bot, which could benefit those who cannot read or write.

1.4 Major Contributions

Given the importance of sexual education for a developing country like India, the proposed work plans to develop and deploy SUKHSANDESH[4], an AI-based QA system for sexual education for the rural Indian population. Several risk factors may arise in the absence of well-executed awareness programs and platforms. Also, there are stigmas about sex-related topics, especially in rural India. Thus, keeping this in mind, we aim to address the following objectives:

1. **A Multi-staged AI-based QA System**: We intend to build a QA system which is robust enough to generate responses for existing and new queries. A retrieval module will be used for the existing queries to retrieve the QA pair, which is most similar to the query. Meanwhile, for new queries, generative AI, such as Large Language Models (LLMs), will be used.

2. **Empathetic Avatar**: We aim to incorporate a 3D avatar that will relay the QA system's response into real-time au-

---

[4]"Sukh" means pleasure, and "Sandesh" means message or communication.

dio and video with humane facial expressions. This will facilitate empathy and engage the users. Furthermore, the audio-based avatar has the potential to attract rural people out of curiosity. Additionally, a significant portion of the rural population lacks the education level needed for reading and writing. This avatar-based QA system will be effective in this case.

3. **Regional Language Support**: India is a diverse country with 22 official languages in the Eighth Schedule[5]. We intend to support regional languages by using various multilingual and openly available translation APIs.

4. **Safety and Trustworthiness**: Given the sensitivity and the stigmas surrounding sex-related topics in India, we intend to provide a safe platform that will be non-judgemental, ensure the user's privacy and build overall trust. We aim to approach this by meticulously using safety protocols such as sanitisation and filtering of private information and employing AI guardrails to prevent harmful and unwanted generations.

5. **Scalable Model**: We also intend to collaborate closely with Gram Vaani, our industry partner, to gather more new data and incrementally improve and scale the overall performance of SUKHSANDESH using online learning from the new data over time.

Overall, our proposal addresses multiple dimensions of sustainable development, including health, education, gender equality, reduced inequalities, and partnership building.

2 Goals

The primary goal of this work is to provide a safe and secure platform to the vulnerable sections of rural India concerning sexual education where the users can ask, learn and become aware of healthy practices without the fear of getting stigmatised. This work revolves around a multi-staged QA system that aims to build a safe and trustworthy AI system supporting regional language.

Multi-staged QA System: At the global level, a multi-staged QA system for sexual education by leveraging both retrieval and generative AI is less explored. There are reports of the usage of chatbots for sexual health in the UK [Nadarzynski *et al.*, 2021], Bangladesh [Rahman *et al.*, 2021], USA [Bonnevie *et al.*, 2021] and South Africa [De Filippo *et al.*, 2022]. However, these systems were solely based on static questionnaires, which lacked empathy and did not consider the contextual information of the overall chat. We intend to exploit the combination of retrieval and LLM-based systems to give responses to even unseen queries based on the contextual information from the database and other information from the user profile for a more personalised and empathetic response. On top of this, we also plan to integrate a 3D real-time avatar to cater for a text-free experience for the users. The avatar will serve two primary purposes: first, to foster empathy through real-time facial and tonal expressions, and second, to provide accessibility to individuals who face challenges with reading or writing, thus

---

[5]https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf

catering to the needs of the rural Indian context.

Safety Guardrails: AI guardrail is a strategy to filter out the model's input and output. A guardrail takes a set of objects as input. These objects can be the input or output of LLMs. The guardrail then determines if any enforcement actions can be taken. Here, the goal is to reduce the risks embedded in these objects [Dong *et al.*, 2024]. Guardrails aim to detect potential misuse during the query stage and work to prevent the model from providing inappropriate answers. AI guardrail is necessary where generative AI is used irrespective of any domain. There is a report [Menz *et al.*, 2024] on using LLMs for spreading misinformation in the medical domain. The authors used LLMs to generate 102 blogs on vaccines and vaping in just 60 minutes. The blogs targeted multiple diverse societal groups. This raises concern regarding the exploitation of LLMs to spread misinformation on such critical issues. Most sexual health-related systems focus on surface-level safety protocols like personal information sanitisation/anonymisation. However, there are other spectrums of safety protocols, such as harmful and unwanted content generation and misinformation. If these protocols are not addressed, it will lead to distrust among the user. Given the sensitivity of sex-related topics in India, it is extremely important to have meticulous and scrutinised guardrails to safeguard the users' privacy and build trustworthiness.

3 Alignment with SDGs

Our proposal is rightly aligned with the following SDGs:

- **Good Health and Well-being (SDG 3):** Sexual education aims to foster a healthy lifestyle, which is crucial for emotional, mental, and social well-being. By addressing issues such as early pregnancy, unsafe abortions, sexually transmitted infections, and sexual violence, sexual education contributes to promoting good health and well-being.

- **Quality Education (SDG 4):** We aim to provide a safe and trustworthy platform for sexual education, especially tailored to the rural Indian population. This initiative aligns with the goal of ensuring inclusive and equitable quality education, as it focuses on delivering essential counseling and information to a vulnerable demographic group.

- **Gender Equality (SDG 5):** Sexual education plays a significant role in promoting gender equality by addressing issues related to sexual and reproductive health, including sexual violence and early pregnancy. By providing access to comprehensive sexual education, especially in regions where adolescents face significant vulnerabilities, this proposal contributes to advancing gender equality.

- **Reduced Inequalities (SDG 10):** In many countries, including India, access to sexual education is unequal, with rural populations often facing barriers due to stigma and lack of resources. By designing a platform tailored to the needs of rural India and ensuring regional language support, the proposal aims to reduce inequalities in access to essential sexual education resources.

- **Partnerships for the Goals (SDG 17):** The proposal involves partnering with Gram Vaani for deployment. This collaboration reflects the importance of partnerships be-

tween governments, civil society, and the private sector in achieving the Sustainable Development Goals.

4 SUKHSANDESH: Model Description

In this section, we elucidate the methodology of SUKHSANDESH for sexual education in rural India. We propose a multifaceted pipeline towards the QA system by employing various deep learning techniques for retrieval and generation tasks. Furthermore, considering the sensitivity and the taboo around sexual education in rural India, the QA system will incorporate AI guardrails [Menz *et al.*, 2024; Dong *et al.*, 2024] against sensitive or unwanted behaviours and mock questions. In this section, we will discuss the dataset and individual modules of SUKHSANDESH. Figure 1 presents a schematic diagram of SUKHSANDESH.

4.1 Dataset Sources

To develop a functional QA system that resonates with the rural Indian demographics, a dataset representing this spectrum is required to make a functional QA system. In this regard, we use a question-answer dataset by Gram Vaani, our industry partner and a non-profit organisation working for rural India on the ground level. The KAB (Kahee Ankahee Baatein) dataset contains QA pairs, in which questions sharing the same answer are grouped to represent various ways of asking the same question. The dataset consists of more than 3900 QA entries; it features more than 1200 unique sanitised questions and 3708 unique caller query transcriptions. It is categorised into broad themes (such as sexual education, safe sex, sexual intercourse, pregnancy, reproduction, puberty, etc.), enabling more efficient answer retrieval when theme information is available. The dataset transcribes a telephonic conversation between the user and the domain expert. The following are some of the features of the dataset:

- **Caller Query**: This field contains the link to the telephonic audio file of the user's question.
- **Caller Query Transcription**: The text version of the telephonic audio query after the transcription.
- **Relevant Question**: Callers often include personal and irrelevant information in their questions. This irrelevant information is removed from these questions to create the relevant question.
- **Sanitised Question**: The question contains some sensitive information sanitised to create a refined version. These sanitised questions can be mapped to the corresponding relevant questions.
- **Theme and Sub-Theme**: The caller tags all the query transcriptions with a corresponding theme and sub-theme.

We propose to leverage the sexual education theme from this dataset for the QA system. The answers will be retrieved for the existing related questions, while new questions will be fed to the LLMs to generate the answers. Meanwhile, we will augment KAB with more human-generated questions and answers received from health workers, which will be incorporated into the model over time.

4.2 Audio-to-Text Conversion of Telephonic Transcripts

The input to SUKHSANDESH will be a user-generated audio query. However, the intermediate QA generation from both the retrieval and generative AI-based approach requires in-text format as these are text-only models. The source of the KAB dataset is the audio recordings from the Interactive Voice Response (IVR) of the users in Hindi. These audio recordings in Hindi are directly translated into English text via Bhasini's audio-to-text translation API [6]. Only the text is required to train and infer the intermediate QA modules. During the inference, initially, the user will give input via audio. This makes the platform accessible to even the underprivileged section of society. This audio query in Hindi will then be translated into English text via the Bhasini API, and the QA modules will generate the response. Finally, the talking head avatar animates the text response into real-time video with emotion. However, most avatar models are in English; therefore, we need to finetune them with Hindi video data. Additionally, the input to SUKHSANDESH is Hindi audio, and the final desirable output from the empathetic avatar is Hindi audio. Therefore, instead of translating the Hindi audio into English text and then feeding the text into an English-only LLM, we plan to simply recognise the Hindi text using Bhasini's automatic speech recognition (ASR) API. This will reduce translation-related errors and maintain the demographic and cultural tones. Furthermore, this Hindi text can be directly fed to Hindi LLMs [Gala *et al.*, 2024] or multilingual LLMs [Scao *et al.*, 2022].

4.3 Retrieve Previous Questions

The KAB dataset, compiled by grassroots NGOs and self-help groups (SHGs), reflects the demographic makeup of rural India. As a result, we anticipate encountering numerous new queries that resemble existing ones. We will leverage these related queries through a retrieval-based system to address this. This approach involves retrieving answers for the most pertinent queries based on their relevance to existing ones. Initially, we will explore both the dense [Luan *et al.*, 2021; Izacard *et al.*, 2021] and sparse retrieval [Robertson *et al.*, 2004] approaches and select the optimal one. As we need to retrieve only the relevant QA pair, it is crucial to quantify the relevance of the retrieved document. Furthermore, we will explore the re-ranking-based approach [Sung *et al.*, 2023; Fajcik *et al.*, 2021] for quantifying the relevance score and come up with some threshold value to consider or discard the retrieved QA pair.

4.4 QA Generation from LLMs

Retrieval-based QA is effective when a similar query is already present and does not account for the client's nuances and preferences. However, LLMs have shown their efficacy in flexible QA generation tasks [Jiang *et al.*, 2021] by considering the contextual information. The LLM generation is typically performed in the following ways. (i) **Finetuning**: The existing QA dataset can be used to finetune the LLM using a supervised finetuning [Ouyang *et al.*, 2022] approach. For

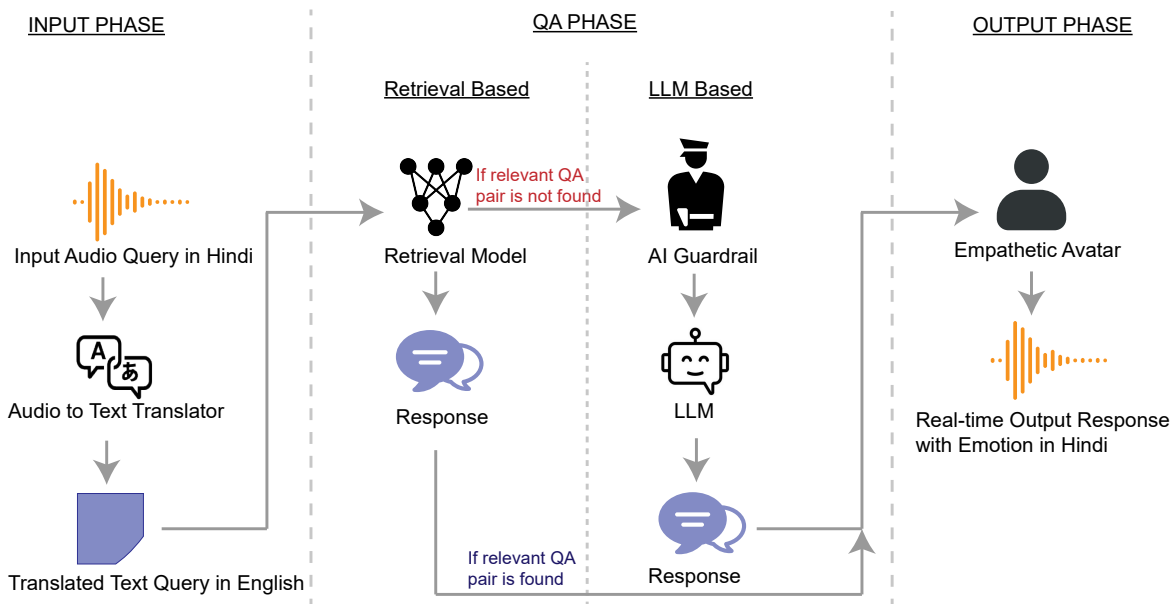---

[6]https://anuvaad.bhashini.gov.in/

Figure 1: A schematic diagram of SUKHSANDESH, our proposed multi-staged QA system for sexual education. SUKHSANDESH consists of three stages, namely, the input phase, the QA phase and the output phase. Initially, it inputs the user's audio query in a regional language, i.e., Hindi. The Hindi audio query is translated to English via the audio-to-text translator API of Bhasini. The text query is then fed into the QA modules. The QA module comprises retrieval-based and LLM-based modules. The retrieval-based module retrieves the most relevant QA pair with the query, and the corresponding answer is returned. LLMs will generate responses with AI guardrails over the query prompt to regularise safety if a relevant pair is not found based on the relevance score. Finally, the text response is converted to real-time audio-visual representation with emotion using a talking head empathetic avatar. It is to be noted that since the output audio is in Hindi, we aim to finetune the existing talking head avatars using Hindi video data.

a decoder-only LLM such as Llama [Touvron *et al.*, 2023], Bloom [Scao *et al.*, 2022], Mistral [Jiang *et al.*, 2023], Open-Hathi[7], and Airavata [Gala *et al.*, 2024], the question and answers are concatenated into a single string, and the model is tasked to generate the next token in an autoregressive manner. Bloom is a multilingual LLm among these LLMs, while OpenHathi and Airavata are LLMs in Hindi. Meanwhile, for an encoder-decoder LLM such as T5 [Raffel *et al.*, 2020], mT0 [Muennighoff *et al.*, 2023] and Aya [Üstün *et al.*, 2024], the question is treated as the source and the answer as the target. T5 is an English-only LLM, while mT0 and Aya are multilingual models supporting Hindi. (ii) **In-context Learning:** This approach does not train or finetune the existing LLM [Dong *et al.*, 2022]. Rather, it exploits the generalisability of the model via prompting with examples of the question-answer pairs to guide the generation of the LLM. We propose utilising LLMs' generative prowess (viz., encoder-decoder and decoder only) for QA generation for sexual education in rural India by exploring fine-tuning and in-context learning approaches.

4.5 Empathetic Avatar

Avatar-based chatbots and systems have been predominantly used as therapeutic modules [Goonesekera and Donkin, 2022; Vaidyam *et al.*, 2019]. In this work, our goal is to use a 3D avatar. *Specifically, we propose to use AI talking head avatar [Weisman and Peña, 2021] which enhances the*

---

[7]https://www.sarvam.ai/blog/announcing-openhathi-series

human-computer interaction. These AI talking head avatar converts text into speech and simultaneously animate the avatar by synthesising lip motion sequences corresponding to the text [Zhen *et al.*, 2023]. The overall output is a real-time video where the avatar seems to talk like a human with visual and tonal emotions according to the context. In our pretext, the avatar will convert the text responses generated by the QA system in real time. These responses can be from either the retrieval-based method or LLMs. The conversion will result in an audio-visual interpretation of the text response. The avatar will try to build empathy by showing real-time, humane facial expressions according to the context. We also explore the opportunity to extend the system further by providing audio-based queries instead of text so that the users won't need to type any text. Moreover, a significant portion of the Indian rural population lacks basic education, hindering them from reading or writing abilities. This text-free avatar-based QA system will be of great potential as it won't require the users to have reading or writing abilities, and their curiosity to use the avatar will increase the overall usage of the QA system for sexual education.

4.6 Support to Regional Languages

One of the major challenges for a QA system in the rural Indian setting is the language barrier. India is a diverse country with multiple regional languages and dialects. A large section of the rural population is not fluent or does not know English. However, most existing LLMs are English-centric and will be

ineffective for regional languages. To address this, we aim to generate the Hindi QA dataset. However, this would require a machine translation (MT) system trained using the regional language. In this regard, we will only focus on Hindi as the choice of the regional language as it is spoken and understood by a majority of the Indian population, and there are already existing powerful MT systems for English-Hindi pairs [Kunchukuttan *et al.*, 2018] and abundant parallel corpus. We also plan to extend this to other popular Indian languages, such as Tamil and Bengali. However, existing MT systems cannot generate sensitive, technical terms like reproductive organs in Hindi. We will also explore the domain-specific perspective of the model. Additionally, instead of utilising an English-centric LLM and postprocessing the QA generation into the regional language, we propose extending the LLM into the regional language for our specific task.

4.7 Guardrails Against Sensitive or Unwanted Behaviours

Current LLMs are extremely powerful, but can be adjusted with prompts. This flexibility of LLMs comes with several risks, but not limited to sensitive data leaks, abusive responses, immoral queries, and misinformation that malign the integrity of the AI system and have a serious impact on users and societies [Dong *et al.*, 2024; Menz *et al.*, 2024]. In the context of our proposal, sexual education remains a taboo subject, leading individuals to shy away from openly expressing their views or engaging in discussions about it. Hence, it is critical to sanitise or anonymise the identity and filter out any personal information in the dataset to increase the trustworthiness and a sense of privacy among the users. Additionally, the LLM-based QA system may generate undesired behaviours such as misinformation and toxic, insensitive, vulgar and harmful content. *We aim to construct guardrails against these potential risks to promote a safe and credible AI system for QA generation.* Particularly, we will finetune the model with some instruction prompts where we will explicitly instruct to generate harmless responses. The initial guardrail will be a meticulously chosen prompt and later on, we will explore additional advanced guardrails.

5 Deployment Plan with Gram Vaani

Existing Capabilities of Gram Vaani. Gram Vaani currently operates the Kahi Ankahi Baatein (KAB) service over an IVR system. A toll-free phone number was publicized via a network of partners working in the areas of gender equality, sexual and reproductive health, grassroots media, etc. Users can give a call to this phone number and cut the call. The IVR server then automatically calls them back within a few seconds. Over this voice call, users can use keypresses to navigate across published audio messages, ask questions by recording a voice message, and listen to answers to questions asked by others. Incoming questions are heard by a team of trained content moderators. The KAB service started in 2015 and receives an average of 300 daily calls from adolescents and young couples. New questions or topical questions are put up to a gynaecologist partner who has been associated with the KAB service for several years; she provides voice-

recorded answers to the questions published on the KAB IVR the following week.

SUKHSANDESH's Deployment Strategy. As presented in this proposal, the KAB service is being enhanced to also operate over a smartphone chatbot and to use the new capabilities of generative AI to improve the interactivity of the service. The service will also be linked via the existing KAB IVR and publicized through two further venues. First, Gram Vaani runs a community volunteer network in more than 1000 panchayats in Northern India (states of Bihar, Jharkhand, Uttar Pradesh, and Madhya Pradesh); the volunteers assist communities with getting access to social entitlements under government schemes, get answers to queries by farmers related to agriculture, learn about digital financial services and online safety, etc. [Moitra *et al.*, 2016]. Over 40% of the volunteers are women. This face-to-face connection by the volunteers will be used to publicize the service further through local youth networks. Second, in these panchayats, Gram Vaani operates an IVR service similar to KAB but with more generic content related to hyperlocal news, health, and social issues such as early marriage and domestic violence etc. Over 100,000 unique users access the service each month, about 30% of whom are youth under 25 years of age [Seth *et al.*, 2020]. We will also publicize the chatbot via this service. Overall, we expect to achieve a usage of more than 1000 users per day of the chatbot service. With regular updates, maintenance, and user feedback, we will ensure that the system remains secure, stable, user-friendly, and continually meets the needs of its users.

6 Model Evaluation

We plan to evaluate the performance of SUKHSANDESH based on multiple factors as mentioned below:

- **Precision and Accuracy**: The model should be able to generate an adequate and fluent answer that the gold labels will gauge. Evaluation metrics such as ROUGE [Lin, 2004] and BERTScore [Zhang *et al.*, 2019] can be utilised.
- **Bias**: The model should be agnostic towards any bias and should be fair and impartial in its generation.
- **Extent of Hallucination**: As a generative AI model, the proposed system may be prone to hallucination. However, the model should only generate adequate responses and reduce hallucinations. Chainpoll [Friel and Sanyal, 2023] can be used as an evaluation metric for this aspect.
- **Generalisability**: The model should be able to generalise well to new data that it has not seen before. For this, we perform any precision and accuracy evaluation metric on the new data.
- **Interpretability**: The model should be interpretable for generation. This will facilitate causal relations between the input features of the model and the overall QA generations.
- **Robustness**: The model should be robust to changes in the data and to any noise or errors that may be present. The proposed method is a pipeline of multiple units; any errors propagated in the channel should not affect the overall performance.
- **Scalability**: The model should be scalable to handle large volumes of data efficiently. And it should be able to learn

incrementally from new data.

7 Challenges

An AI-based QA system like SUKHSANDESH poses multiple challenges. These challenges get multiplied when applied to a rural Indian pretext. The following are some of the challenges of a QA system for sexual education in rural India:

- **Taboos and Stigmas Related to Sex**: A large number of the rural population in India is stigmatised from openly speaking and sharing their view on sex-related topics, specifically females. This hinders the data acquisition step, where the users are expected to share and ask educational questions related to sex. Furthermore, even if the model is deployed, then also the stigma will prevail with using a sexual education tool. Moreover, many people normalise or do not consider the problems as problems such as having more than four children or not sending the girl child to school. Therefore, a substantial effort at the ground level for awareness is required, which we expect to overcome with Gram Vaani's ground-level workers
- **Lack of Labeled Data**: LLMs necessitate a significant volume of data for training to achieve satisfactory performance. However, acquiring sex education-related data, particularly within the rural Indian context, is challenging due to the taboo surrounding topics related to sexuality. Consequently, gathering sufficient data poses a major challenge. Moreover, SUKHSANDESH will require machine translation or alternative multilingual strategies to accommodate regional languages. Nevertheless, the availability of such multilingual resources is also constrained.
- **Data Quality**: Sex education-related data can be subject to quality issues such as gender bias, normalising taboos, or misaligned instances in the parallel corpus for the machine translation phase, leading to biased results, or the data may not be representative of the population being studied. And, as our pipeline requires speech-to-text conversion, there might be cases where the error in the conversion might get propagated further downstream.
- **High Carbon Footprint**: LLMs require significant energy, contributing to carbon footprints. This impact can adversely affect Sustainable Development Goal 7 (Affordable and Clean Energy) and Sustainable Development Goal 13 (Climate Action).
- **Lack of Skilled Personnel**: AI model training and deployment require skilled manpower. Additionally, these personnel must be adaptive to the growing pace of AI. Nonetheless, there is a need for more skilled AI personnel in India.

8 Risks, Limitations, Ethical Considerations

The taboos and stigmas w.r.t. sex-related topics prevail among the rural Indian population. Until and unless ground-level awareness is carried out, it will be a limitation and hurdle in the overall project. Furthermore, the use of LLMs comes with the issue of hallucination. We aim to address this by forcing the LLMs to generate only factually correct responses; however, the model may not be fully hallucination-proof. We aim to use a 3D avatar for real-time text-to-speech interpretation with emotions. However, perfect emotions with facial structure will be difficult to mimic. Furthermore, if done wrongly, it will have an uncanny behaviour which might be uncomfortable sometimes. We also plan to incorporate AI safety guardrails into the model to protect and filter out any unwanted and harmful content. However, it is important to note that these guardrails may be tweaked because designing robust guardrails against any new adversaries is challenging. Finally, there might be some ethical considerations in data acquisition and their use, particularly sensitive user information. This will require their consent and sanitisation of the data to filter out and anonymise their information.

9 Expected Results and Long-Term Plans

It is expected that the comprehensive findings of the planned study on sexual education in rural India will offer a deep understanding of the complex issues surrounding sex. The project intends to address societal biases towards sex by identifying the factors that contribute to the stigmas associated with sexual education. Adopting a customised, private quality assurance system is anticipated to increase target users' trustworthiness and confidence. Long-term plans also call for creating a personalised education ecosystem. This will meet the changing needs of the Indian rural population. Furthermore, community-based engagement and awareness programmes will make a secure platform for sexual education available. Lastly, the integrated results can offer different perspectives on spreading awareness by utilising target profiles.

10 Conclusion

Sexual-related issues constitute a significant concern in rural India. People are often unaware of sexual education due to the stigmas and taboos prevailing in sex. The subject is not openly discussed, and those who attempt to address it are often met with social disapproval. This raises a pressing concern, which is evident with the normalisation of child marriage, unprotected sex, unsafe abortions, lack of menstrual sanitation and maternity care. To address these issues, our proposal aims to provide a safe and private space for this vulnerable and marginalised section of the Indian population. Specifically, we propose building SUKHSANDESH, a QA system for sexual education in the rural Indian setting by leveraging the prowess of traditional AI and generative models. Moreover, for enhanced accessibility and user convenience, we also plan to use a real-time avatar-based QA system. This system will prioritise empathy, where the input query will be audio-based in Indian languages. The responses generated will be the real-time audio-visual representation of the text response from the QA system. The system will be scalable and continuously adaptive. SUKHSANDESH will incorporate new training data and get updated over time. Additionally, to expand the outreach, SUKHSANDESH will also support regional languages. Finally, we will also integrate AI guardrails for security and privacy concerns to prevent harmful and unwanted responses. In collaboration with Gram Vaani, we will deploy SUKHSANDESH in rural India as a mobile-based application to spread widespread awareness of sexual education, which is still an abandoned and controversial topic in India.

Acknowledgments

References

[Agarwal *et al.*, 2021] D Agarwal, A Agastya, M Chaudhury, T Dube, B Jha, P Khare, and N Raghu. Measuring effectiveness of chatbot to improve attitudes towards gender issues in underserved adolescent children in india. Technical report, Tech. Rep.). Cambridge, MA: Harvard Kennedy School, 2021.

[Bonnevie *et al.*, 2021] Erika Bonnevie, Tiffany D Lloyd, Sarah D Rosenberg, Kara Williams, Jaclyn Goldbarg, and Joe Smyser. Layla's got you: Developing a tailored contraception chatbot for black and hispanic young women. *Health Education Journal*, 80(4):413–424, 2021.

[Datta and Majumder, 2012] Shib Sekhar Datta and Nilratan Majumder. Sex education in the school and college curricula: Need of the hour. *J Clin Diagn Res*, pages 1362–4, 2012.

[David *et al.*, 2013] Miriam David, S Walker, and L Barton. Sex, education and social policy: a new moral economy? In *Gender, Class and Education (Routledge Revivals)*, pages 141–159. Routledge, 2013.

[De Filippo *et al.*, 2022] Alexandra De Filippo, Paloma Bellatin, Neville Tietz, Alexander Whitefield, Puseletso Nkopane, Eli Grant, Kaitlyn Crawford, Camilla Devereux, Benjamin Vermeulen, and Abigail Hatcher. Effects of digital chatbot on gender beliefs and exposure to intimate partner violence among young women in south africa. 2022.

[Dong *et al.*, 2022] Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, and Zhifang Sui. A survey for in-context learning. *arXiv preprint arXiv:2301.00234*, 2022.

[Dong *et al.*, 2024] Yi Dong, Ronghui Mu, Gaojie Jin, Yi Qi, Jinwei Hu, Xingyu Zhao, Jie Meng, Wenjie Ruan, and Xiaowei Huang. Building guardrails for large language models. *arXiv preprint arXiv:2402.01822*, 2024.

[Fajcik *et al.*, 2021] Martin Fajcik, Martin Docekal, Karel Ondrej, and Pavel Smrz. R2-D2: A modular baseline for open-domain question answering. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *EMNLP*, pages 854–870, November 2021.

[Friel and Sanyal, 2023] Robert Friel and Atindriyo Sanyal. Chainpoll: A high efficacy method for llm hallucination detection. *arXiv preprint arXiv:2310.18344*, 2023.

[Gala *et al.*, 2024] Jay Gala, Thanmay Jayakumar, Jaavid Aktar Husain, Mohammed Safi Ur Rahman Khan, Diptesh Kanojia, Ratish Puduppully, Mitesh M Khapra, Raj Dabre, Rudra Murthy, Anoop Kunchukuttan, et al. Airavata: Introducing hindi instruction-tuned llm. *arXiv preprint arXiv:2401.15006*, 2024.

[Ghule and Donta, 2008] Mohan Ghule and Balaiah Donta. Sexual behaviour of rural college youth in maharashtra, india: An intervention study. *Journal of Reproduction and Contraception*, 19(3):167–189, 2008.

[Goonesekera and Donkin, 2022] Yenushka Goonesekera and Liesje Donkin. A cognitive behavioral therapy chatbot (otis) for health anxiety management: Mixed methods pilot study. *JMIR Form Res*, 6(10):e37877, Oct 2022.

[Holohan and Fiske, 2021] Michael Holohan and Amelia Fiske. "like i'm talking to a real person": Exploring the meaning of transference for the use and design of ai-based applications in psychotherapy. *Frontiers in Psychology*, 12:720476, 2021.

[Ismail *et al.*, 2015] Shajahan Ismail, Ashika Shajahan, T S Sathyanarayana Rao, and Kevan Wylie. Adolescent sex education in india: Current perspectives. *Indian J Psychiatry*, 57(4):333–337, October 2015.

[Izacard *et al.*, 2021] Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. Unsupervised dense information retrieval with contrastive learning. *arXiv preprint arXiv:2112.09118*, 2021.

[Jiang *et al.*, 2021] Zhengbao Jiang, Jun Araki, Haibo Ding, and Graham Neubig. How Can We Know When Language Models Know? On the Calibration of Language Models for Question Answering. *TACL*, 9:962–977, 2021.

[Jiang *et al.*, 2023] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[Khubchandani *et al.*, 2014] Jagdish Khubchandani, Jeffrey Clark, and Raman Kumar. Beyond controversies: sexuality education for adolescents in india. *J Family Med Prim Care*, 3(3):175–179, July 2014.

[Kumar *et al.*, 2017] Randhir Kumar, Anmol Goyal, Parmal Singh, Anu Bhardwaj, Anshu Mittal, and Sachin Singh Yadav. Knowledge attitude and perception of sex education among school going adolescents in ambala district, haryana, india: A Cross-Sectional study. *J Clin Diagn Res*, 11(3):LC01–LC04, March 2017.

[Kunchukuttan *et al.*, 2018] Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. The IIT Bombay English-Hindi parallel corpus. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

[Lin, 2004] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. ACL.

[Luan *et al.*, 2021] Yi Luan, Jacob Eisenstein, Kristina Toutanova, and Michael Collins. Sparse, Dense, and Attentional Representations for Text Retrieval. *TACL*, 9:329–345, 2021.

[Maeda *et al.*, 2020] Eri Maeda, Akane Miyata, Jacky Boivin, Kyoko Nomura, Yukiyo Kumazawa, Hiromitsu Shirasawa, Hidekazu Saito, and Yukihiro Terada. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. *Reproductive BioMedicine Online*, 41(6):1133–1143, 2020.

[Mehrolia *et al.*, 2023] Sangeeta Mehrolia, A. Veena, and Subburaj Alagarsamy. Inclusion of sexual health education for the well-being and dignity of secondary school children: An indian rural perspective. *Journal of Health Management*, 25(4):792–801, 2023.

[Menz *et al.*, 2024] Bradley D. Menz, Natansh D. Modi, Michael J. Sorich, and Ashley M. Hopkins. Health Disinformation Use

Case Highlighting the Urgent Need for Artificial Intelligence Vigilance: Weapons of Mass Disinformation. *JAMA Internal Medicine*, 184(1):92–96, 01 2024.

[Mills *et al.*, 2023] Rhiana Mills, Emily Rose Mangone, Neal Lesh, Diwakar Mohan, and Paula Baraitser. Chatbots to improve sexual and reproductive health: Realist synthesis. *J Med Internet Res*, 25:e46761, Aug 2023.

[Moitra *et al.*, 2016] Aparna Moitra, Vishnupriya Das, Gram Vaani, Archna Kumar, and Aaditeshwar Seth. Design lessons from creating a mobile-based community media platform in rural india. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, pages 1–11, 2016.

[Muennighoff *et al.*, 2023] Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. Crosslingual generalization through multitask finetuning. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *ACL*, pages 15991–16111, July 2023.

[Nadarzynski *et al.*, 2021] Tom Nadarzynski, Vannesa Puentes, Izabela Pawlak, Tania Mendes, Ian Montgomery, Jake Bayley, and Damien Ridge. Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis. *Sex Health*, 18(5):385–393, November 2021.

[Ouyang *et al.*, 2022] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *NeurIPS*, volume 35, pages 27730–27744, 2022.

[Raffel *et al.*, 2020] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), jan 2020.

[Rahman *et al.*, 2021] Rifat Rahman, Md. Rishadur Rahman, Nafis Irtiza Tripto, Mohammed Eunus Ali, Sajid Hasan Apon, and Rifat Shahriyar. Adolescentbot: Understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in bangladesh. 2021.

[Ray *et al.*, 2012] Subha Ray, Shailendra Kumar Mishra, and Bhubon Mohan Das. Sexual and reproductive health issues among rural and urban adolescent boys of eastern india. *Journal of Men's Health*, 9(2):94–101, 2012.

[Robertson *et al.*, 2004] Stephen Robertson, Hugo Zaragoza, and Michael Taylor. Simple bm25 extension to multiple weighted fields. In *CIKM*, page 42–49, New York, NY, USA, 2004. Association for Computing Machinery.

[Scao *et al.*, 2022] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, et al. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.

[Seth *et al.*, 2020] Aaditeshwar Seth, Akshay Gupta, Aparna Moitra, Deepak Kumar, Dipanjan Chakraborty, Lamuel Enoch, Orlanda Ruthven, Paramita Panjal, Rafi Ahmad Siddiqi, Rohit Singh, et al. Reflections from practical experiences of managing participatory media platforms for development. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, pages 1–15, 2020.

[Sung *et al.*, 2023] Mujeen Sung, Jungsoo Park, Jaewoo Kang, Danqi Chen, and Jinhyuk Lee. Optimizing test-time query representations for dense retrieval. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5731–5746, Toronto, Canada, July 2023. Association for Computational Linguistics.

[Touvron *et al.*, 2023] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[Tripathi and Sekher, 2013] Niharika Tripathi and T V Sekher. Youth in india ready for sex education? emerging evidence from national surveys. *PLoS One*, 8(8):e71584, August 2013.

[Üstün *et al.*, 2024] Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, et al. Aya model: An instruction finetuned open-access multilingual language model. *arXiv preprint arXiv:2402.07827*, 2024.

[Vaidyam *et al.*, 2019] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7):456–464, 2019. PMID: 30897957.

[Wang *et al.*, 2022] Hua Wang, Sneha Gupta, Arvind Singhal, Poonam Muttreja, Sanghamitra Singh, Poorva Sharma, and Alice Piterova. An artificial intelligence chatbot for young people's sexual and reproductive health in india (snehai): Instrumental case study. *J Med Internet Res*, 24(1):e29969, Jan 2022.

[Weisman and Peña, 2021] William D Weisman and Jorge F Peña. Face the uncanny: the effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions. *Cyberpsychology, behavior, and social networking*, 24(3):182–187, 2021.

[Zhang *et al.*, 2019] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*, 2019.

[Zhen *et al.*, 2023] Rui Zhen, Wenchao Song, Qiang He, Juan Cao, Lei Shi, and Jia Luo. Human-computer interaction system: A survey of talking-head generation. *Electronics*, 12(1):218, 2023.