AI based Learning App for Students with Special Needs

V. Manoj Prabhakaran, Masters in Computer Science, Georgia Institute of Technology, Atlanta, Georgia, US.

Dr. P. Pandia Rajammal, Assistant Professor, Department of English
Kalasalingam Academy of Research of Education, Tamil Nadu, India.

V. Alex Allan Raja, B.E Computer Science and Engineering
Thiagarajar College of Engineering, Tamil Nadu, India.

ABSTRACT:
This paper discusses the development track of the mobile application "FlexTutor" which is designed to help users of diverse backgrounds and needs to use digital libraries easily and effectively learn using mobile applications. The paper is subdivided into sections that address the status quo, work done and how the application can potentially help to address the issues and the possible improvements to shorten the gaps in effective education systems due to language, accessibility and physical barriers.
Keywords: Mobile application, digital libraries, language, accessibility and physical barriers

INTRODUCTION

Education is often considered as the great equalizer in society to help people live a good quality of life. However, education in its own self, is not equally distributed throughout the world. There are differences in the quality of education systems between different countries in the world such as, accessibility of education, differences in infrastructure and employability, quality of educational boards, coalitions and partnerships with other developed countries, etc. Also, there are differences in the quality of education even within countries such as, inclusion for disabled students, differences in infrastructure within institutions, communal differences, language differences, etc. It is evident that technology has a clear impact on education and it looks promising to improve the quality of education not only in developing countries but also everywhere. Education is the greatest equalizer and technology, in my opinion, is the greatest enabler of the greatest equalizer.

The overall direction so far in developing nations has looked good and it gets more and more promising as time progresses. As discussed in the related works, there are applications for inclusive learning but most of them are frameworks at this point. There is no unified application that helps users of diverse backgrounds and abilities to access educational content in a mobile learning environment when it comes to learning management systems. The integration of LLMs and Language APIs are also on the rise but they are far from achieving inclusive learning features. In developing nations, the divisive factors are larger than usual and hence bridging gaps such as language and disability in learning will be useful.

RELATED WORKS

On the topic of inclusive education for students with special needs, the policies that were introduced by the Indian government (as an example of developing nations) so far on inclusion in education to help children with special needs are good but are not up to the global standards. The paper also does a comparison of how policies are at the global stage (Gulyani, 2017). The changes in policies across all the years and how it has improved inclusion in the nation is described in the paper. The results concluded from another study, (Das, Kuyini & Desai, 2013) said that the majority of the teachers

involved had little to no experience in teaching to students with special needs and there is also a high need of better awareness and faculty training to make the teaching process more inclusive and the paper suggests working on better policies. Apart from the economic and physical challenges, language barriers are also key influencers in the education system of developing nations. As noted in the Paraguay study, (Grazzi & Vergara, 2012) language impacts learning in developing countries. Paraguay is taken for the case study, it has 40% of the population is monolingual in Guaraní, 50% bilingual and 6% monolingual in Spanish. A variety of variables are considered for the study and the distribution is sampled as random as possible. The paper concludes that Guarani is a crucial factor in ICT diffusion. This implies that language is a significant component in the development of e-learning along with other socio-economic factors such as income and level of development, etc. Therefore, we can conclude that language is also a crucial component in designing inclusive learning.

Traditional learning requires everyone to be physically present at the institution. However, due to the increasingly computerized nature of education, it makes it difficult for people with blind and visual impairments to fully use the system. These difficulties include accessing information, evaluation information, locating information, difficulties in searching etc. Studies argue that there is a need for restructuring digital libraries to better help blind and visually impaired people (Xie et al., 2020). Just like visually impaired people, hearing impaired people also face challenges in online mode of education. Adding captions can be one way to make video content more inclusive to people with hearing impairments (Hong et al., 2011). It is shown that the more dynamic and integrated the captions are the better it is for people to follow and use. A study on youtube captioning strategies (Li, Lu, Lu, Carrington & Truong, 2022) also strengthen this fact. Not having proper transcripts or captions in video content proves to be a disadvantage for people with hearing impairments. Adding smart search mechanisms such as similarity scores (Liu, Carrington, Chen & Pavel, 2021) are promising for people with visual difficulties. From these examples, it is clear that significant strides are taken forward in improving inclusion in education in developing nations and technology looks highly promising to solve some of the tough problems in education.

Significant developments have been done in instilling human-level quality responses when converting text input to speech output (Tan, 2024). Hybrid models that combine the use of generative AI to achieve dynamic controllable outputs to speech results (Guo, 2023) were also discussed. Emotional temperature elements were introduced in Emodiff models (Guo, 2023) that help to get close to human-like emotions in speech systems that can be used to convert educational content to speech. One aspect that was interesting but may not be applicable in all areas is the use of non-binary voice based agents (Danielescu, 2023). As a part of developing more inclusive educational systems, Virtuoso (Saeki, 2023) and Maestro-U (Chen, 2023) are helpful in giving design frameworks and considerations when applying multilingual language models and transfer learning across languages while using large language models. The study of the use of speech to text technology in high school (Levine, 2023), provided a knowledge of practical challenges in adopting newer assisted methods of learning in classical and augmented environments, although this is not entirely relevant to inclusive technology, it gave an insight on challenges that students face if they have an accent/dialect that is different from the common use. Furthermore, involving paralinguistic data in language models (Luangrath, 2023) is promising as it can add rich non textual data along with text as inputs. Another promising topic in inclusive AI based technology is the dynamic query based video content management systems (Huang, 2020). This can help get information from videos by giving queries to encoder-decoders. Video summarization strategies are also looking more and more feasible (Hsu, 2023), (Zhang, 2023), (Bodi, 2021). Although there are a lot of limitations in the state of the art video summarization AI, the overall

pace that the field is moving is very rapid. Apart from research based models, commercial models such as GTTS, GSTT and Video-ChatGPT look extremely good and show great performance on benchmark data. They can be used for commercial applications that require high performance, little prototyping time and high quality results.

PROPOSED WORK

The proposed work will be in the developmental track. The app is called FlexTutor that uses Text-to-speech, Speech-to-text, Multi-language translation models, automatic video summarization techniques and large language models to augment users by providing them effective means of learning and accessing educational content in their own comfortable way they prefer.

This would be an Android application that can process videos, audio and video content to help the users and assist them in their learning process. The idea is to either create a python backend in the server when the app calls for processing or to have the models baked into the application itself as an alternate approach.

TECH STACK

- Python, Flask for backend
- Huggingface models for
  - Speech to text (openai/whisper-large-v3)
  - language translation (facebook/seamless-m4t-v2-large)
  - Gemini for LLM
- Android for Front end

TASK LIST

Week #	Task #	Task Description	Estimated Time (Hours)
8	1	Check in with Mentor to finalize work	1
8	2	Initial Model selection for Text to Speech	2
8	3	Initial Model selection for Speech to Text	2
8	4	Initial Model Selection for Language Translation	2
8	5	Initial Model Selection for Video Transcription	2
8	6	Initial Model Selection for LLMS	2
9	7	Proof of Concept for TTS - Initial Model	2
9	8	Proof of Concept for STT - Initial Model	2
9	9	Proof of Concept for Language Translation - Initial Model	4
9	10	Proof of Concept for Video Transcription - Initial Model	4
9	11	Proof of Concept for LLM usage	4
10	12	Evaluation and Improvement Plans for TTS	2

10	13	Evaluation and Improvement Plans for STT	2
10	14	Evaluation and Improvement Plans for Language Translation	2
10	15	Evaluation and Improvement Plans for Video Transcription	2
10	16	Evaluation and Improvement Plans for LLMs	2
INTERMEDIATE MILESTONE 1 DUE			
11	17	Wave 1 of improvements for TTS	2
11	18	Wave 1 of improvements for STT	2
11	19	Wave 1 of improvements for Language Translation	2
11	20	Wave 1 of improvements for Video Transcriptions	2
11	21	Wave 1 of improvements for LLMs	2
12	22	Android Platform integration of TTS model	2
12	23	Android Platform integration of STT model	2
12	24	Android Platform integration of Language Translation	2
12	25	Android Platform integration of Video Transcriptions	2
12	26	Android Platform integration of LLMs	2
13	27	Full deployments and prototype	4
13	28	Testing, Evaluation, Feedback and next steps	4
13	29	Project milestone preparation, fixes and changes	2
INTERMEDIATE MILESTONE 2 DUE			
14	30	Wave 2 of improvements for TTS	2
14	31	Wave 2 of improvements for STT	2
14	32	Wave 2 of improvements for Language Translation	2
14	33	Wave 2 of improvements for Video Transcriptions	2
14	34	Wave 2 of improvements for LLMs	2
15	35	Change Deployment, Integration steps	3
15	36	End user testing, feedback	5
15	37	Final Wave of improvements for TTS	2
15	38	Final Wave of improvements for STT	2
15	39	Final Wave of improvements for Language Translation	2
16	40	Final Wave of improvements for Video Transcriptions	2
16	41	Final Wave of improvements for LLMs	2

16	42	Documentation	4
16	43	Final Evaluations and Feedback	4
16	44	Final Project Preparation for delivery and demo	3
FINAL PROJECT DUE			
Total Hours			**106**

USER SCENARIOS

Some of the user scenarios are discussed below.

- ***User wants to look at content*** - This can be done using the home screen by navigating between Text, Audio and Video content screens
- ***User wants to open a content*** - After navigation, the user can simply click on one of the entries in the UI for the content. A new screen should open up that is specific for the content
- ***User needs to get transcripts*** - The user can click on the view transcripts button for audio and video. For text, the text is already displayed
- ***User needs to read out text content*** - The user can click on the read original button to start playing audio of the text content
- ***User needs to play video content*** - The user can click on play button to open a new video activity screen
- ***User needs to ask questions*** - The tutor button can be clicked and it opens the tutor activity. The user can then type in the question and click on ask. A reply will be generated quickly.
- ***User wants a different language explanation*** - Click on translate button

CONCLUSION AND REFLECTIONS

Looking at the big picture of creating a mobile digital library for people and starting out on using multiple quadrants of tech stacks to synergistically work together was personally satisfying work to me. The application is solid and is built in a scalable fashion with proper api driven architecture. The application is also flexible enough to accommodate intricate changes such as change in underlying models, adding new functionalities, being able to use both CPU and GPU in the backend, based on availability. In terms of usage, it looks promising to bridge the gaps in language and physical barriers when learning. The app being a mobile application is highly useful in adoption and quick access thus helping in better permeance among groups that may not be able to afford PCs. With the use of AI, there were a list of concerns such as privacy and exploitation. Proper prompt techniques and processing paradigms such as building on-server models helped a lot to address the issues. Overall, the development work done is continuous and agile, with regular modular increments in functionality being deployed and tested. I am confident that the work can be extended to real life scenarios and can be helpful given the right circumstances of use.

FUTURE WORK AND IMPROVEMENTS

- The app currently is set up to translate to Tamil language but with the model in the backend, it is possible to translate to a lot more languages. Hence more translation options can be added
- Better UI with cleaner elements
- Try out with new ML models constantly and better hyperparameter optimization, because AI pops up new models and optimizations every second
- Since Android is open source and personally was comfortable given my expertise, Android was chosen

as the scope. However this app can be extended to Flutter and iOS as well

- Have voice based controls

APP UI

Figure 1 *– Home Menu*

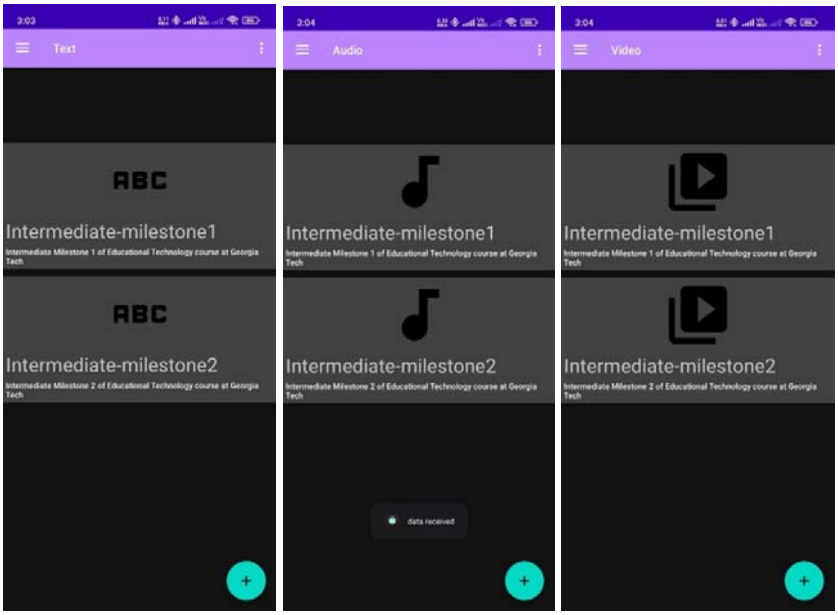

Figure 2 *– Text, Audio and Video content screens*

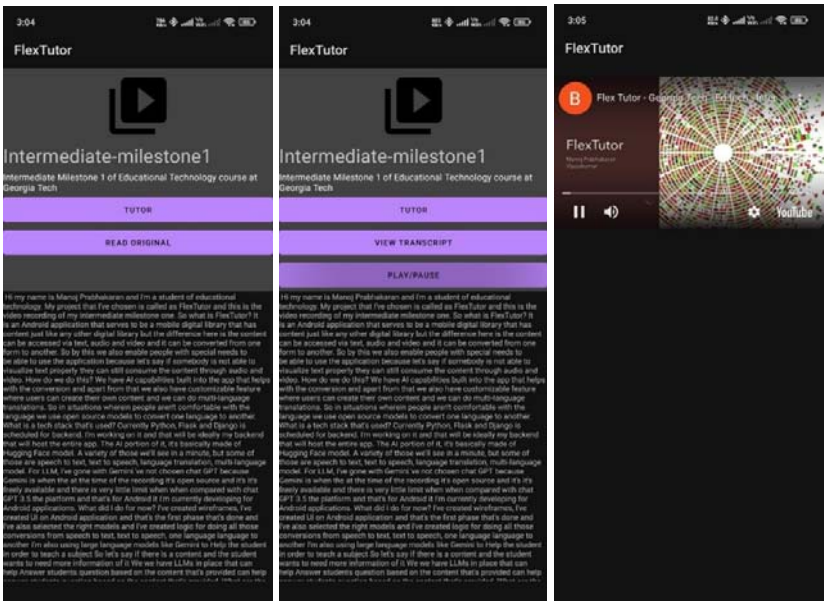

Figure 3 *– Content screens for Text (left), Audio and Video (middle), Video player (right)*

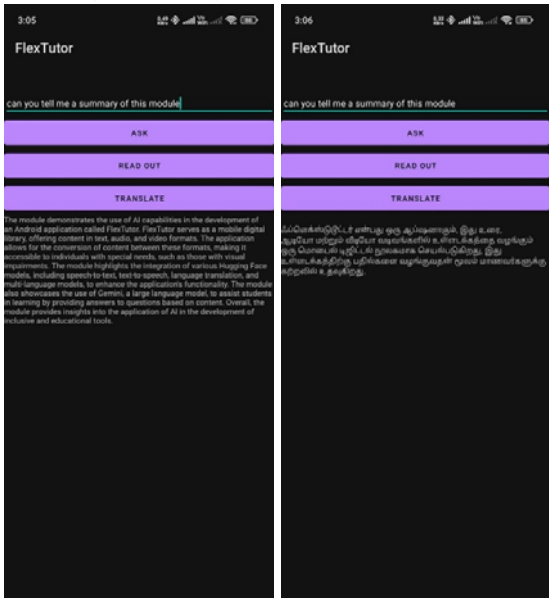

Figure 4 *– Tutor LLM screen (left), Translation (right)*

REFERENCES

1. Khare, M. (2014). Employment, employability and higher education in India: The missing links. Higher Education for the Future, 1(1), 39-62.
2. Gupta, D., & Gupta, N. (2012). Higher education in India: structure, statistics and challenges. Journal of education and Practice, 3(2).
3. Gulyani, R. (2017). Educational policies in India with special reference to children with disabilities.

Indian Anthropologist, 47(2), 35-51.

4. Das, A. K., Kuyini, A. B., & Desai, I. P. (2013). Inclusive Education in India: Are the Teachers Prepared?. International journal of special education, 28(1), 27-36.
5. Forlin, C. (2013). Changing paradigms and future directions for implementing inclusive education in developing countries. Asian Journal of Inclusive Education, 1(2), 19-31.
6. Grazzi, M., & Vergara, S. (2012). ICT in developing countries: Are language barriers relevant? Evidence from Paraguay. Information Economics and Policy, 24(2), 161-171.
7. Tadesse, S., & Muluye, W. (2020). The impact of COVID-19 pandemic on education system in developing countries: a review. Open Journal of Social Sciences, 8(10), 159-170.
8. Kanwal, F., & Rehman, M. (2017). Factors affecting e-learning adoption in developing countries–empirical evidence from Pakistan's higher education sector. Ieee Access, 5, 10968-10978.
9. Khan, H., & Bashar, O. K. (2016). Does globalization create a 'level playing field through outsourcing and brain drain in the global economy?. The Journal of Developing Areas, 50(6), 191-207.
10. Xie, I., Babu, R., Lee, T. H., Castillo, M. D., You, S., & Hanlon, A. M. (2020). Enhancing usability of digital libraries: Designing help features to support blind and visually impaired users. Information Processing & Management, 57(3), 102110.
11. Hong, R., Wang, M., Yuan, X. T., Xu, M., Jiang, J., Yan, S., & Chua, T. S. (2011). Video accessibility enhancement for hearing-impaired sers. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 7(1), 1-19.
12. Li, F. M., Lu, C., Lu, Z., Carrington, P., & Truong, K. N. (2022). An exploration of captioning practices and challenges of individual content creators on YouTube for people with hearing impairments. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1-26.
13. Liu, X., Carrington, P., Chen, X. A., & Pavel, A. (2021, May). What makes videos accessible to blind and visually impaired people?. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
14. Tan, X., Chen, J., Liu, H., Cong, J., Zhang, C., Liu, Y., ... & Liu, T. Y. (2024). NaturalSpeech: End-to-End Text-to-Speech Synthesis with Human-Level Quality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
15. Guo, Z., Leng, Y., Wu, Y., Zhao, S., & Tan, X. (2023, June). PromptTTS: Controllable text-to-speech with text descriptions. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
16. Guo, Y., Du, C., Chen, X., & Yu, K. (2023, June). Emodiff: Intensity controllable emotional text-to-speech with soft-label guidance. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
17. Danielescu, A., Horowit-Hendler, S. A., Pabst, A., Stewart, K. M., Gallo, E. M., & Aylett, M. P. (2023, April). Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).
18. Saeki, T., Zen, H., Chen, Z., Morioka, N., Wang, G., Zhang, Y., ... & Ramabhadran, B. (2023, June). Virtuoso: Massive Multilingual Speech-Text Joint Semi-Supervised Learning for Text-to-Speech. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
19. Chen, Z., Bapna, A., Rosenberg, A., Zhang, Y., Ramabhadran, B., Moreno, P., & Chen, N. (2023, January). Maestro-U: Leveraging joint speech-text representation learning for zero supervised speech

asr. In *2022 IEEE Spoken Language Technology Workshop (SLT)* (pp. 68-75). IEEE.

20. Levine, S., Hsieh, H., Southerton, E., & Silverman, R. (2023). How high school students used speech-to-text as a composition tool. *Computers and Composition*, *68*, 102775.
21. Luangrath, A. W., Xu, Y., & Wang, T. (2023). Paralanguage classifier (PARA): An algorithm for automatic coding of paralinguistic nonverbal parts of speech in text. *Journal of Marketing Research*, *60*(2), 388-408.
22. Huang, J. H., & Worring, M. (2020, June). Query-controllable video summarization. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (pp. 242-250).
23. Hsu, T. C., Liao, Y. S., & Huang, C. R. (2023). Video Summarization With Spatiotemporal Vision Transformer. *IEEE Transactions on Image Processing*.
24. Zhang, H., Li, X., & Bing, L. (2023). Video-llama: An instruction-tuned audio-visual language model for video understanding. *arXiv preprint arXiv:2306.02858*.
25. Bodi, A., Fazli, P., Ihorn, S., Siu, Y. T., Scott, A. T., Narins, L., ... & Yoon, I. (2021, May). Automated Video Description for Blind and Low Vision Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7).