

Refactor the chatbot and voice in the Speak Activity to use gen-AI

~ *Mentors: Ibiam Chihurumnaya & Walter Bender*

Basic Details:

Full Name: **Mebin J Thattil**

Email: **mail@mebin.in; mebin.j.thattil@gmail.com**

GitHub Username: **mebinthattil**

First Language: **English**

Location: **Bangalore, Karnataka, India**

Time zone: **GMT+5:30 (Indian Standard Time)**

Nickname / Preferred name: **Mebin**

Availability: At least **20 hours a week**. I have **no other long-term commitments** during the summer.

Project Length: **175 hours**

Hours Online: **8:00 AM – 12:00 AM IST**

Few personal projects that use python & AI:

- <https://github.com/mebinthattil/Python-Monthly-Finances> - (Python + Llama 3.1 Ollama)
- <https://github.com/mebinthattil/College-Buddy-Junior-Study> - (Ed-Tech project; peer-to-peer learning platform; designed for under-privileged kids)

Synopsis:

This project aims to **modernize the Speak Activity** by **integrating a large language model** into the chatbot feature to enhance its educational value for early learners. The Speak Activity is used for helping beginners engage with reading through synthetic speech.

This project aims to implement an **LLM chatbot** to replace the pre-existing chatbot and to **replace the robotic-sounding synthesized voice** to a better more **natural sounding TTS** model that gets the **pronunciation and phonetics right**. By tuning the LLM to **handle invented spelling** and accommodate **incorrect grammar**, the project will be able to deliver a much better learning experience than the methods used before. Additionally, the chatbot will be designed with a supportive persona, modeling interactions where an adult explains things to a young child. The model aims to both **answer questions** in a simplified manner, as well as **correct the learner** of any misspelt words or grammatical errors.

This project will use **Python** and a **fine-tuned model** hosted on **AWS**, which is accessed via an **API endpoint**. This ensures that once the project is completed and tested, it can be merged with the pre-existing speak activity **without any additional dependencies** and is fully compatible.

Note: The previous chatbot and voice will not be removed from the speak activity. Those shall be used as a fallback if the user does not have a stable internet connection, or if the user chooses to opt-out of AI services.

Benefits:

Benefits To Speak Activity:

- The chatbot can take the persona of a parent/teacher trying to explain and guide the learner, thus making the learning experience a little bit **more personalized**.
- We can **avoid instances** of speak AI **not working** because of some kind of **environment issues**. As the new speak AI would just make API calls to the model hosted on AWS, this would ensure that it works independent of what environment you have locally. [Link to github issue.](#)
- Better understands made-up spellings.
- Contextual-aware chat, better follow-up questions.

Benefits To Sugar:

The model developed during this project can **serve as the foundation** for many other AI-enhanced activities. The fine-tuned model will have qualities such as comprehension of **fabricated spellings** and **appropriate filters and safeguards for children**. Hence this fine-tuned model can be **adapted for other sugar activities** with minimal modifications.

E.G: This same model can be used in the finance activity to give insights on spendings. OR. This model can be used in turtle blocks as a general chatbot to ask questions about that activity.

Benefits To Open-Source Community:

- Since the code and models are open source, anyone can change the model that the speak activity uses. This can be good if someone wants to try the speak activity with a model of their choice or wants to run a model locally and try it as well. Switching models will be made very easy.

Deliverables:

Goals:

- A new fine-tuned model with the following requirements:
 - Tight guard-rails and age-appropriate filters [**High Priority**]
 - Light-weight (1.5-2B parameters) [**optional** might go above / below]
 - Must accommodate made up spellings [**required**]
 - A friendly persona that guides the learner [**required**]
 - Contextual recommendations and adapt to the learner [**required**]
 - Follow up questions to yield better learning outcomes (e.g. “Have you tried typing this word out?”) [**required**]

Note: Sugar-AI will be used/tuned if that’s what’s preferred over fine-tuning another completely different model. (depends on the mentor’s view on this)

- Integrate a TTS model with the following requirements:
 - sounds more natural and less robotic. [**required**]

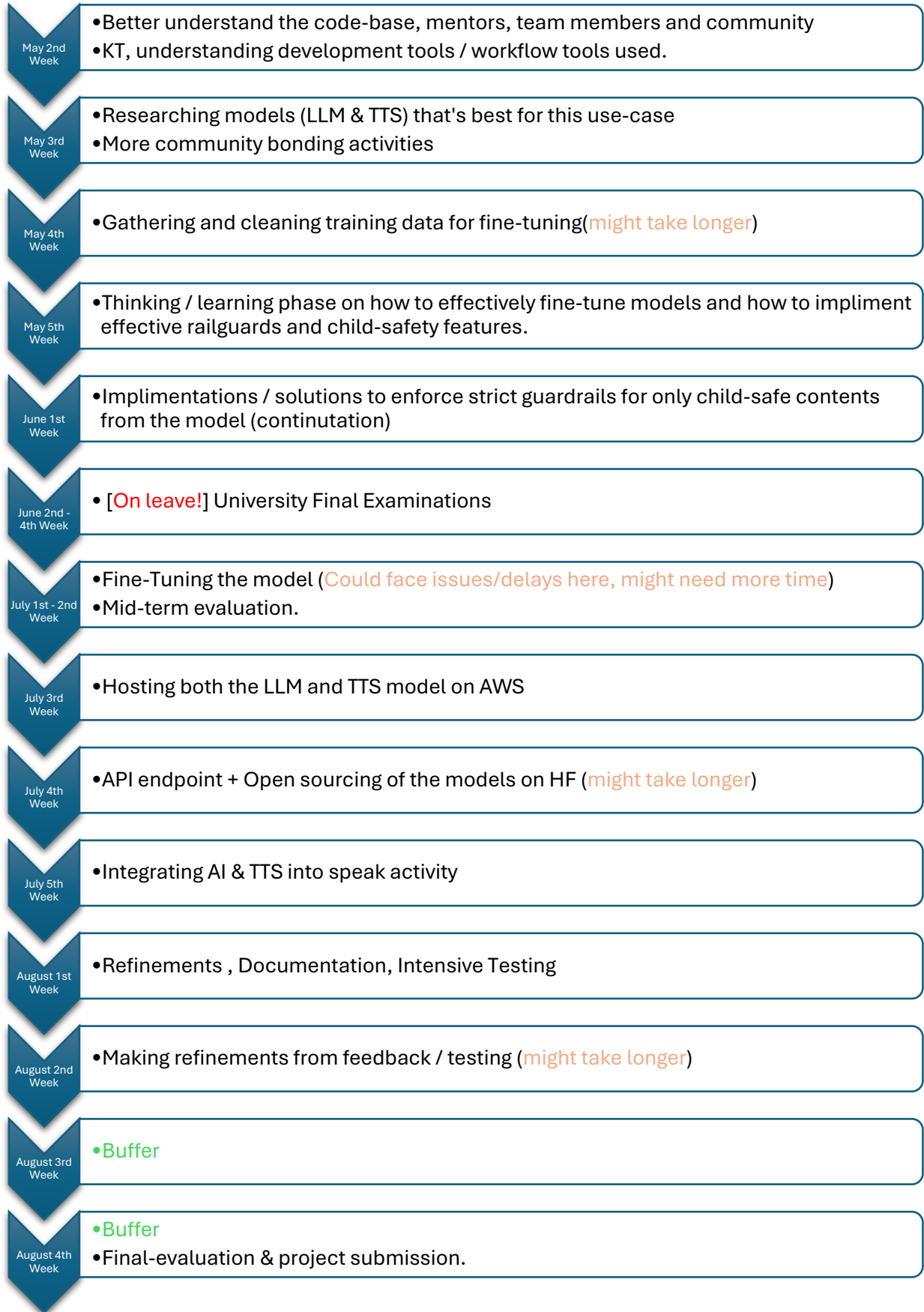
- gets pronunciation and phonetics right [**required – Higher Priority**]
- Hosting on AWS and creating an API endpoint for the fine-tuned LLM and TTS model, so that it can be easily integrated into speak activity as well as other sugar activities if required in the future.. [**required**]
- Documentation – adding as well as improving pre-existing documentation. [**required**]
- Enhancing UI – current UI is outdated. Expect improvements in accessibility and intuitiveness. [**required – but lower priority**]
- Not over-achieving. Instead, creating a project that’s **safe for kids** and thoroughly tested. By the end of summer, it must be merged to the speak activity. [**High Priority**]

Milestones:

- Researching and finalizing on an open-source model for our use case.
- Gathering and cleaning data to fine-tune the model
- Fine-tuning the model (*I plan on using unsloth for this*)
- Hosting the model on AWS and creating an API endpoint
- Open sourcing the model on HF
- Integrating the model into speak activity
- Finding best model for TTS
- Hosting the TTS model on AWS and creating API endpoints
- Integrating the new TTS voice to speak activity
- Open sourcing the TTS model on HF
- Testing, Documentations, merging to main
- Additional [**Optional – lower priority**]:
 - Optimizing model for speed
 - Explore options to optimize cloud & hosting costs

Timeline / Deadlines:

Note: Since I have university exams during **June 7th – June 27th**, I have included some coding work under the community bonding time period. I will be on leave during my exams.



Why Sugar should choose me:

1. I have a fair understanding of the speak codebase and have tinkered around with a few things.
2. I have a small demo where I used a pre-existing model from Hugging Face to power the chatbot in the speak activity. Demo can be found here: [\[link\]](#).
Doing this has given me the confidence that I can work on this project and complete it during the summer.
3. My development environment is all setup, so I can get up and running much quicker.
4. Although I'm in my first year of college, I have been coding for almost 7 years now, and this is a golden opportunity that would let me enhance my coding skills further and push me a little deeper into the world of open-source development.
5. I have worked on a previous personal project that used both generative AI and python. Working on the speak activity would help broaden my knowledge and sharpen my AI and python development skills.

Related Work:

Last year, sugar integrated AI into its activities, enhancing their functionality. E.g.:

<https://summerofcode.withgoogle.com/archive/2024/projects/lccFY8AH>

As part of my project, I not only aim to integrate AI with speak, but also hopefully lay a solid foundation for AI that can be used in many other sugar activities.

Bio:

Hello! I am Mebin J Thattil, a first-year undergraduate student studying Computer Science and Engineering at PES University, Bangalore, India. I have a deep passion for coding and tech in general. I've been fascinated by the world of computer science from the age of 10. Over the years I have slowly gotten to know more about the subject, while building cool things along the way. I also enjoy playing video games and badminton.

I'm a strong advocate for technology in education and have built a web-application that enables students in rural school of India to connect with high school students in urban cities to take courses for free (peer-to-peer free course portals) [\[link\]](#)

I have also conducted various CS related workshops for junior grades, so I have a fair idea of working with kids and I understand how they learn and interact with technology.

Recently I've been developing my personal portfolio website, that is hosted on a raspberry pi3 running at home. I learnt and implemented technologies like network tunneling, domain registrars, WSGI, NGINX, CDN/S3 buckets, etc to build the infrastructure. - mebin.in (it's still WIP & entire source-code should be open-sourced within a few weeks).

I feel like I'm a good candidate to undertake this project as I have worked on previous personal projects that have the same technical requirements this one. I have also tried a very basic version of AI implementation on the pre-existing speak activity, this has given me the confidence that I can complete this project, meeting expectations and would hopefully be able to merge by the end of the summer.

I am a newcomer to bigger FOSS projects in general but have collaborated with various people on campus on smaller open-source projects and initiatives. I believe in giving back to the community. I'm also part of the technical team of the largest FOSS club in my campus where we spread awareness about FOSS and conduct workshops and other events like hacknight (hacktoberfest).

Checkout my GitHub for more POW - <https://github.com/mebinthattil>

Checkout my portfolio site - <https://mebin.in/>