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**

Portfolio Website: mebin.in

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)

[Also see: my demo integrating Qwen 2.5 model into speak activity's chatbot: LINK]

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 (Large Language Model) chatbot** to replace the pre-existing chatbot and to **replace the robotic-sounding synthesized voice** to a better more **natural sounding TTS (Text To Speech)** 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 **answer questions** in a simplified manner, **correct the learner** of any misspelt words or grammatical errors, and **guide** the learner on **how to learn** in a better way (e.g.: explaining methods to learn algebra in a simple way; or, asking the learner to try and pronounce a certain word and try to spell it).

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.

This project will also include packaging a small and lightweight model along with the speak activity (something like BERT / DistilBERT) and fine-tune that as well. This will be used as **a fall back** if the learner **does not have an active internet connection**. This ensures that the user would still be able to enjoy the benefits of an LLM based system, with the only caveat being slightly worse responses than the model hosted on AWS.

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.
- Better more natural sounding voice.
- UI enhancements – loading indicators, improved intuitiveness with features like hints and highlighted elements while using for the first time.

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 LLM model (hosted on AWS) 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)

- A new fine-tuned LLM model (packaged with speak) with the following requirements:
  - Ultra-light weight (something like Distil BERT, exact model to be used will be finalized later.) [required]
  - All the aforementioned features and guardrails as the same as the model hosted on AWS. [required]

- 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. Plan on implementing features like loading indicators, highlighted elements that guide the learn where to click on screen (while using for the first

time), a system that provides users with hints on how to use the speak activity more efficiently (e.g. pointing out that things like pitch can be changed if the user has never tried it in a long time.)[required – but lower priority]

- Not over-achieving (*For e.g.: I will prioritize refining a single high-quality voice instead of adding multiple voice options, ensuring clarity and usability*).
  I will always prioritize creating something that's **safe for kids**, helps **achieve the learning outcomes**, and is **thoroughly tested**. [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 Hugging Face (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
- Fine-tuning the light-weight model packaged into speak (*plan on using unsloth for this*)
- Integrating the model into speak activity and ensure compatibility.
- 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.

May 2nd Week	• Better understand the code-base, mentors, team members and community • Knowledge Transfer, understanding development tools / workflow tools used.
May 3rd Week	• Researching models (LLM & TTS) that's best for this use-case • More community bonding activities
May 4th Week	• Gathering and cleaning training data for fine-tuning(might take longer)
May 5th Week	• Thinking / learning phase on how to effectively fine-tune models and how to imlpement effective railguards and child-safety features.
June 1st Week	• Implementations / solutions to enforce strict guardrails for only child-safe contents from the model (continutation)
June 2nd -4th Week	• [On leave!] University Final Examinations
July 1st - 2nd Week	• Fine-Tuning the model + build a training pipeline (can be used for the smaller model as well.) (Could face issues/delays here, might need more time) • Mid-term evaluation.
July 3rd Week	• Hosting both the LLM and TTS model on AWS
July 4th Week	• API endpoint + Open sourcing of the models on HF (might take longer)
July 5th Week	• Integrating AI & TTS into speak activity
August 1st Week	• Traning the light-weight fall back model packaged with speak. (might extend to the next week as well) • Integrate this model with the activity.
August 2nd Week	• Refinements , Documentation, Intensive Testing
August 3rd Week	• Making refinements from feedback / testing (might take longer)
August 4th Week	• Buffer • Final-evaluation & project submission.

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. I have a fair understand and have worked with cloud services such as AWS and GCP (google cloud) before.
5. Although I'm in my first year of college, I have been coding for almost 7 years now, and I see this as 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.
6. 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 and meeting expectations.

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/