

## EE/CprE/SE 491 WEEKLY REPORT 8

11/1/2024 – 11/7/2024

**Group number:** 35

**Project title:** Universal Response Engine: LLMs for Good

**Client &/Advisor:** Ahmed Nazar and Mohamed Selim

### **Team Members/Role:**

Abraham Toutoungi - Stakeholder Liaison

Gabriel Carlson - Communications Manager

Halle Northway - Meeting Coordinator

Brianna Norman - Project Deliverables Manager

Ellery Sabado - Timeline Coordinator

Emma Zatkalik - Assignment Manager

---

### Weekly Summary

We made a lot of progress with ETG and our VM. We made a unified model and trained it on the VM. We are still working on improving it and getting datasets that work well. We also made our first commit to GitLab and made a basic frontend that we can use for the demo.

### Past Week accomplishments

- VM up and running with GUI
- Put a unified model on the VM
- Made our first commit to GitLab
- Frontend for the demo is made

### Pending Issues

- N/A

### Individual Contributions

Name	Individual Contributions	Hours this week	Hours cumulative
Abraham Toutoungi	<ul style="list-style-type: none"><li>- Worked on getting the basic functionalities of the frontend</li><li>- Played around on the vm</li><li>- Got gitlab setup</li><li>- Working on connecting frontend to backend</li></ul>	6	44
Gabriel Carlson	<ul style="list-style-type: none"><li>- Worked on transferring google colab</li></ul>	7	44

	<ul style="list-style-type: none"> <li>- unified LLM to VM</li> <li>- Separated out LLM training and LLM model loading into different files</li> <li>- Configured virtual environment for needed python3 dependencies</li> <li>- Debugged using LLama3.2 8B model</li> <li>- Switched implementation to LLama3.2 3B model for VRAM issues and switched to locally saved models</li> <li>- Started collecting performance metrics for training models</li> </ul>		
Halle Northway	<ul style="list-style-type: none"> <li>- Worked on setting up VM and transferring work to VM <ul style="list-style-type: none"> <li>- Downloaded Visual Studio Code</li> <li>- Organizing desktop</li> <li>- Resolving dependency issues</li> <li>- Cleaning up existing fine-tuning code</li> <li>- Modularizing and debugging separate fine-tuning python scripts, one using Unsloth and one using Auto classes from HuggingFace</li> </ul> </li> </ul>	6	45
Brianna Norman	<ul style="list-style-type: none"> <li>- Worked on VM setup for project</li> <li>- Researched Cybersecurity needs for project</li> <li>- Looked into user protection for resource searching varying on state</li> <li>- Researched how the project is affected by HIPPA laws</li> </ul>	6	44
Ellery Sabado	<ul style="list-style-type: none"> <li>- Worked on more fine-tuning with different models</li> <li>- Added more to the dataset and cleaned it up</li> <li>- Researched more on the Weights and Biases graphs for my FineTune model</li> <li>- Looked around in the VM</li> </ul>	6	44
Emma Zatkalik	<ul style="list-style-type: none"> <li>- Got VM running on my computer</li> <li>- Recorded time comparisons for RAG LLM using llama3.1 8B and llama3.2 3B</li> <li>- Finetuning continuation</li> <li>- Looking at datasets</li> </ul>	6	45

Comments and extended discussion (optional)

N/A

#### Plans for upcoming week

- Work on training the model better
- Gather a couple datasets that we can train the model with for the demo
- Connect the frontend and backend for demo

#### Summary of weekly advisor meeting

- If an LLM returns a resource, look at where the resource is used. LLM should cite sources for RAG
  - Industry panel: How do you know what the LLM is giving is true or credible
- Create a mini server for VM for models so we aren't running multiple at the same time
- Find cases that you like in existing data sets, copy them, and merge them with other data sets
- Tensorboard or wandb
  - Take what is happening internally and output to the graph for fine-tuned
- **Work on getting an LLM fully fine-tuned on one dataset**
  
- Week of 18th of November
  - Work on presentation
- Week of 2nd
  - Start practicing presenting
- Week of 9th
  - Presentation