

## **EE/CprE/SE 492 WEEKLY REPORT 3**

**Sept 20 - Oct 3**

**Group number: 20**

**Project title: GridGPT**

**Client & Advisor: Gelli Ravikumar**

**Team Members/Role:**

- **Tin Ngo -> AI Integration Specialist/Backend Lead**
- **Jackson Phillips -> AI Integration Specialist**
- **Emma Heithoff -> Power Systems Specialist**
- **Eddy Andrade -> Frontend Lead**
- **Nick Doty -> Power Systems Specialist**

### ○ **Weekly Summary**

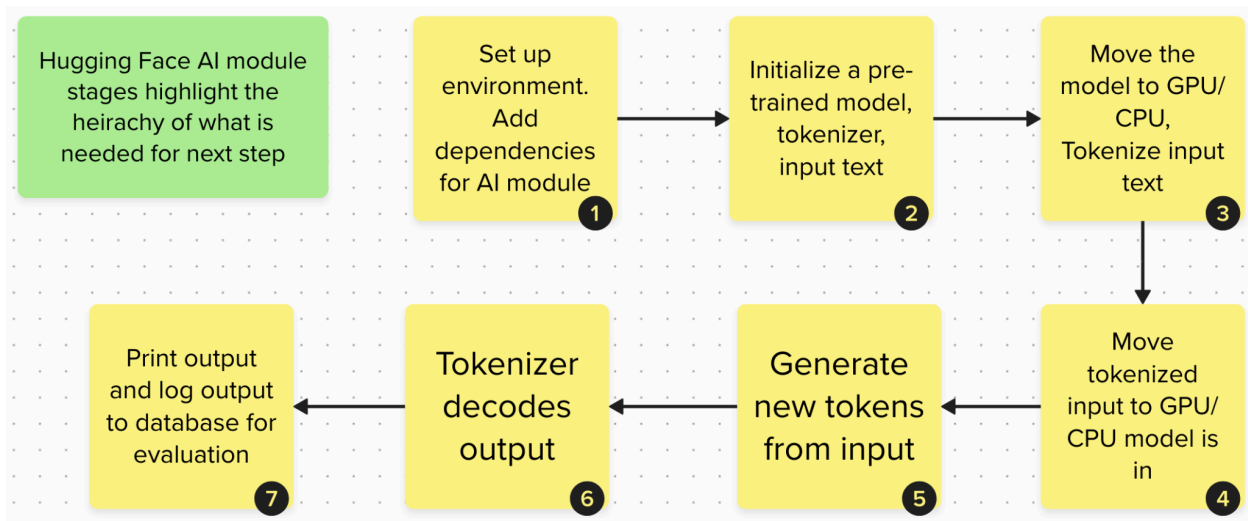
Our three project teams — AI, Frontend, and Power Systems (grid data) — spent time working together in order to collaborate on gaps in our respective implementations. Simulations to gather power grid data were started through troubleshooting from the AI team. Helpful documentation for frontend questions was provided by the graduate students affiliated with the project. The AI team made progress in preparing for eventual model training on the natural language prompt and response dataset from grid simulation.

### ○ **Past week accomplishments**

- Tin Ngo: **Week 1**

I did more research on Hugging Face and grabbed a really small but well-trained model off of Hugging Face named 'appvoid/palmer-002.5'. I created a hugging face module within our project. I made a class that initializes a HuggingFace model, tokenizer, and config. This sets up the LLM locally, and you can send a prompt that it can use to generate text, which is its method. Currently, it takes the CPU cores for power, otherwise, when it is on GCP, it will use cuda, which is GPUs. The inputs are tokenized by the tokenizer and sent to the CPU cores for processing. The model then generates outputs that are returned. It follows

this diagram.



I then created a Flask app that you can send a POST request to, to use this class that generates a text. You simply include the model you want to use, the prompt, and the maximum length of the output.

[https://git.ece.iastate.edu/sd/sddec24-20/-/merge\\_requests/9/diffs](https://git.ece.iastate.edu/sd/sddec24-20/-/merge_requests/9/diffs) is the PR

Here is me testing the new code changes. You can see it successfully generates the output to create a python method that adds two numbers together:

```
(venv) tinngo@Tins-MacBook-Pro sddec24-20 % curl -X POST http://127.0.0.1:5000/generate_text \
-H "Content-Type: application/json" \
-d '{"prompt": "Generate a python method to add two numbers", "max_length": 100}'
Generate a python method to add two numbers

def add_numbers(num1, num2):
    """
    This method adds two numbers and returns the result.
    """
    return num1 + num2

# Example usage
num1 = 5
num2 = 7
result = add_numbers(num1, num2)
print(f"The sum of {num1} and {num2} is: {result}")
```

## Week 2

To get started with HPC, I made a quick Hello World Python script and ran it to make sure I could upload runnable code. I worked with Emma Heithoff to create a script that would run the dss simulation using altdss. Here is that script and the generated output:

[https://git.ece.iastate.edu/sd/sddec24-20/-/merge\\_requests/12](https://git.ece.iastate.edu/sd/sddec24-20/-/merge_requests/12)

Worked on getting a Flask app on HPC to see if it works and it works just as planned!

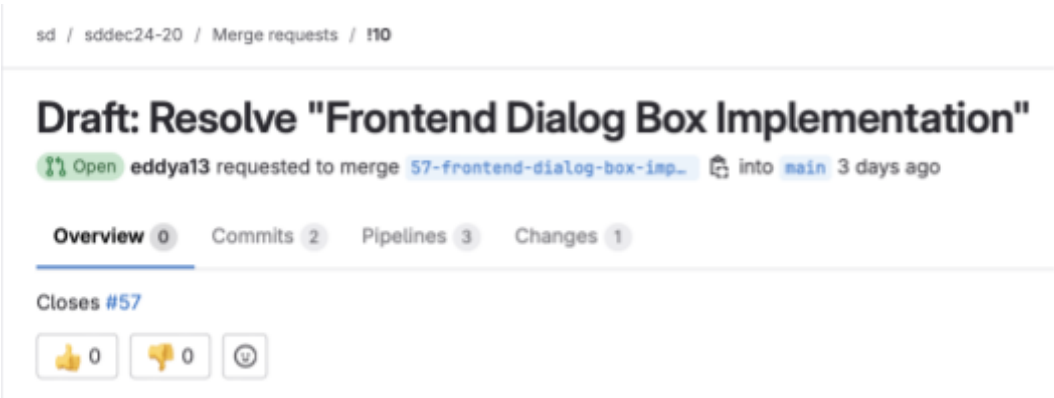
```
Dockerfile __init__.py README.md src tests venv
(venv) [tinngo56@nova huggingFace]$ ls
Dockerfile __init__.py README.md src tests venv
(venv) [tinngo56@nova huggingFace]$ curl -X POST http://127.0.0.1:5000/generate_text \
  -H "Content-Type: application/json" \
  -d '{"prompt": "Generate a python method to add two numbers", "max_length": 100}'
Generate a python method to add two numbers

def add_numbers(num1, num2):
    """
    This method adds two numbers and returns the result.
    """
    return num1 + num2

# Example usage
num1 = 5
num2 = 7
result = add_numbers(num1, num2)
print(f"The sum of {num1} and {num2} is: {result}")
(venv) [tinngo56@nova huggingFace]$ curl -X POST http://127.0.0.1:5000/generate_text \
  -H "Content-Type: application/json" \
  -d '{"prompt": "Generate a python method to add two numbers", "max_length": 100}'

WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
config.json: 100%|#####| 717/717 [00:00<00:00, 107kB/s]
model.safetensors.index.json: 100%|#####| 15.7k/15.7k [00:00<00:00, 2.57MB/s]
model-000001-of-000001.safetensors: 100%|#####| 2.20G/2.20G [00:20<00:00, 108MB/s]
Downloading shards: 100%|#####| 1/1 [00:21<00:00, 21.46s/it]
tokenizer_config.json: 100%|#####| 1.34k/1.34k [00:00<00:00, 170kB/s]
tokenizer.model: 100%|#####| 500k/500k [00:00<00:00, 94.3MB/s]
tokenizer.json: 100%|#####| 1.84M/1.84M [00:00<00:00, 13.9MB/s]
special_tokens_map.json: 100%|#####| 551/551 [00:00<00:00, 298kB/s]
The attention mask and the pad token id were not set. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
Setting 'pad_token_id' to 'eos_token_id':2 for open-end generation.
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
127.0.0.1 - - [02/Oct/2024 12:59:19] "POST /generate_text HTTP/1.1" 200 -
The attention mask and the pad token id were not set. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
Setting 'pad_token_id' to 'eos_token_id':2 for open-end generation.
```

- Jackson Phillips: I got started with HPC and learned how to copy and run code on it. Also got started writing code for training.
- Eddy Andrade: I spoke with Peeyush, the former Frontend developer for the project and asked him for some resources to help me progress through the project. He told me that we're using the Mantine library for components, and suggested I implement the Dialog window and modify it to work with our project, which is what I plan on doing. Finally, I began working on implementing the project by branching out of the main branch of the project, and created the folder ChatComponent and file ChatWindow.tsx.



*Screenshot showing the branch created*

```

1 + import {useClient} from "react";
2 + import {useDisclosure} from "@mantine/hooks";
3 + import {Dialog, Group, Button, TextInput, Text} from "@mantine/core";
4 +
5 + function testDemo() {
6 +   const [opened, {toggle, close}] = useDisclosure(false);
7 +
8 +   return (
9 +     <>
10 +       <Group justify="right">
11 +         <Button onClick={toggle}>GridGPT</Button>
12 +       </Group>
13 +
14 +       <Dialog opened={opened} withCloseButton onClose={close} size="lg" radius="md">
15 +         <Text size="sm" mb="xs" fw={500}>
16 +           Welcome to GridGPT!
17 +         </Text>
18 +
19 +         <Group align="flex-end">
20 +           <TextInput placeholder="Message GridGPT" style={{ flex: 1 }} />
21 +           <Button onClick={close} size="compact-md"></Button>
22 +         </Group>
23 +       </Dialog>
24 +     </>
25 +   )
26 + }

```

Screenshot showing the implementation of the component

**Dialog**

Display a fixed overlay dialog at any side of the screen

Import: `import { Dialog } from '@mantine/core';`

Source: [View source code](#)

Docs: [Edit this page](#)

Package: [@mantine/core](#)

Documentation | Props | Styles API

**Usage**

`Dialog` is a simplified version of `Modal` component. It does not include most of accessibility and usability `Modal` features:

- Focus trap is not available
- Does not close on click outside
- Does not have overlay

Use `Dialog` to attract attention with not important information or action, for example, you can create subscription form:

Clicking this button pops up the Dialog window

Toggle dialog

Dialog window example

Subscribe to email newsletter

hello@gueticker.com

Subscribe

Screenshot showing documentation and the component being used

- Emma Heithoff: I clarified my misunderstanding of building the dataset in our advisor meeting. I ran this simulation to generate line-line and line-neutral voltages by bus and node, circuit element currents, and controlled transformer tap settings.

```

import altdss as dss

# Access the prime_api_util instance
api_util = dss.prime_api_util

# Create the AltDSS simulator
simulator = dss.AltDSS(api_util)

# Create a new DSS context
context = simulator.NewContext()

# Load the Run file which will invoke the master file
run_file_path = "C:/Users/Emma/Desktop/Important/Run_IEEE123Bus.DSS" # Adjust
the path if necessary
context.Load(run_file_path) # Load the run file

# Run the power flow simulation
context.Solution.Solve() # Execute the power flow analysis

```

- Nick Doty: Unfortunately I have not had too much time this week but I have continued to refresh myself with DSS research and messing around with the VM

## **Week 2**

I was able to fix my VM issues by using the VM virtual keyboard to access the shift key.

Additionally, the team was able to meet this past weekend and I have been reading through a chat document between Emma and chatgpt about DSS files and what each line means in the Run\_IEEE123Bus.DSS file and how to perform simulations with DSS files using ALTDSS-python.

## ○ **Pending issues**

- Tin Ngo: Instead of working on hosting an LLM on HPC, I found out that we are actually only supposed to do training on the HPC clusters. We actually want to host our LLM on GCP which is something I do not have access to currently.
- Jackson Phillips: Currently, I have to do all my work on my personal computer instead of the VM because the VM doesn't connect to HPC with the standard ssh command I use on my pc.
- Emma Heithoff: No pending issues.
- Eddy Andrade: No pending issues.

- Nick Doty: N/A

○ **Individual contributions**

<b><u>NAME</u></b>	<b><u>Individual Contributions</u></b> <i>(Quick list of contributions. This should be short.)</i>	<b><u>Hours this week</u></b>	<b><u>HOURS cumulative</u></b>
Tin Ngo	Hosting app on HPC, and working on dss	14	115
Jackson Phillips	HPC learning	6	81
Emma Heithoff	Dataset clarification	12	81
Eddy Andrade	Mantine documentation reading; Frontend Implementation	8	80
Nick Doty	Continued DSS refresher	6	84

○ **Plans for the upcoming week**

- Tin Ngo: Work with Professor Gelli to get access to GCP and host a small LLM on GCP. I am unfamiliar with the services that GCP provides so I will have to do some research. Work with HPC team to try and deploy LLM to HPC. Work with Rolf to deploy LLM to GCP.
- Jackson Phillips: I am going to continue working on code and learning about how to train.
- Emma Heithoff: I will bring a small part of the dataset to the design team meeting for collaboration before continuing my work on the data. Nick and I will brainstorm milestones for our dataset role to work cohesively to interface with the AI training team.
- Eddy Andrade: Plan to work on implementing something (as simple as a button) into the project. Read up on how to implement a chat window.
- Nick Doty: I plan to continue to meet with Emma to plan out what needs to be done next and additionally meet up with Tin and Jackson with Emma to coordinate the input of DSS files into the AI model.

○ **Summary of weekly advisor meeting**

This week, we agreed to build project milestones between the three project teams. Additionally, we clarified goals for our alpha model when we wrap this course. We shared our progress and questions from our workday last Saturday and have information to work in areas where we had questions.