Dear faculty members and Students

I want to pose a few AI challenges to all the students / faculty members or a mixed team of DSU. I don't believe that any particular branch has a monopoly on working on AI techniques. Note that the solutions exist on the net. Enough material is available on YouTube, Github, discussion sites, and technical papers. You just have to implement it. Needless to say, this is the life skill you must have to get a job in the near future and win prizes in competitions. You may also use it to enhance your research skills.

1. Learn to generate a large amount of "synthetic data" from any given data. This would be a helpful skill when large enough data is unavailable for training or the data is guarded by the secrecy clause (eg. patient data). This data can be used for the following challenges.

2. Design a Retrieval Augmented Generation (RAG) system. You may want to download Facebook's open-source Llama 3.1 8B model (don't go for the 180B model, which would be too large) or any other free-to-download LLM. You will have to download the Ollama app. You may use the "Langchain" framework to integrate Llama 3.1 with your RAG. It can be done on a reasonably powerful desktop with i7 processor and NVidia 4GB graphics card. It may require 50-60 GB of free HD space

3. Design a "Small Language Model' (SML)

4. Fine Tune your LLM to do a focused search

5. Create word embedding in the LLM

You should be able to explain your logic and code. First five teams (maximum five members in each team - students, faculty members or a combination) will be recognized by the DSU. Please do the work relate to this challenge in your free time. Don't let your regular studies suffer.

All the best

Amit Bhatt