# ADTA 5900/5770.501: Generative AI with LLMs

## Thuan L Nguyen, PhD

## Assignment 5

## 1. Overview

The rise of cloud computing has facilitated the emergence of big data. Cloud computing commodifies computing time and data storage using standardized technologies.

Big data is a term for large volumes of data that can be both structured and unstructured. These enormous volumes of data overwhelm the digital world every second. However, it is not the amount of data that matters. What we can do with the data matters: Big data analytics can provide insights that lead to better decisions and strategic moves.

The emergence of cloud computing has made it easier to provide the best technology in the most cost-effective packages. Cloud computing has reduced costs and made many applications available to companies of all sizes: small, medium, large, and giant corporations.

## 2. Google Cloud Platform (GCP): Service Account

Various enterprise services of Google Cloud Platform (GCP) will be used in classwork. When using cloud services to run applications, cyber-security is one of the top priority. To create and run Natural Language Processing (NLP) applications in GCP, the user must set up a service account that can be used in GCP's sophisticated authentication system.

The project uses the Google Cloud Platform (GCP) Vertex AI service. Each student must have a GCP account so that he/she can access and use GCP: Vertex AI services as required.

**IMPORTANT NOTES**:
*--) All the documents posted on the Canvas page **GOOGLE CLOUD PLATFORM: GCP for Deep Learning – TF2** can be used for HW 4, HW 5, and the final project.*

*--) All the documents posted on the Canvas page **GOOGLE CLOUD PLATFORM: GCP for Natural Language Processing (NLP)** can be used for HW 4, HW 5, and the final project.*

## 3. Homework 5: Assignment Format

Homework 4 is assigned as a team assignment. It means all the group's student members will collaborate while working on the assignment.

However, each student must **write** and **submit** his/her report **independently**. In other words, a student works on the assignment with the team but **writes** and **submit**s the report as if he/she had worked on the assignment independently.

## 4. Homework 5: General Assignments

Each group is assumed to be an AI system development team in a business organization. With the explosion of popularity and widespread use of generative AI in real-world management and business activities, the corporation's leaders want the team to develop a generative AI system that the company employees can use to perform content searches, ask questions, and get answers about the contents of the organization's proprietary documents.

The team will adopt Google Cloud Platform (GCP): Vertex AI services as the primary system Integrated Development Environment (IDE) to design, build, and test the system throughout the project, including but not limited to cloud storage, vector embeddings generation, vector databases management, and advanced vector search technologies. For development, the group will use Python for coding with Google Collaboratory (Colab) as the coding IDE. The group also plans to use popular generative AI techniques, including but not limited to Retrieval Augmented Generation (RAG), Sentence Transformer, and tools provided by generative AI platforms like LangChain and Hugging Face.

## 5. PART I: Generative AI Q&A-Search System: System Analysis (50 Points)

The student is required to submit a project analysis report as the solution to Homework 5.
The report must include the following sections:

1. **Introduction**
   a. Provide an overview of the project.

2. **Problem statement**
   a. Discuss in detail what problem (business, technical, …) the student is trying to solve with this project.

3. **System Requirements Analysis (HW 4)**
   a. Business requirements
      i. Discuss in detail the business requirements of the generative AI system.

   b. Technical requirements
      i. Discuss in detail the technical requirements of the generative AI system.

   c. Data requirements
      i. Discuss in detail the data requirements of the generative AI system.

4. **Feasibility Analysis**
   a. Technical feasibility analysis:
      i. Can we complete the project successfully as required?
      ii. Discuss any technical risks while working on the project.

   b. Business feasibility analysis:
      i. Will the project provide good business value after its completion?
      ii. Discuss any financial risks while working on the project, e.g., running out of funding
   c. Operation feasibility analysis
      i. If we build the system, will it be used by the organization as expected?
      ii. Discuss any risks that may hinder the system's deployment after its completion.

5. **Project Management**
   a. Discuss in detail the timeline of the project
      i. What kinds of significant tasks are to be done?
      ii. What major phases need to be done until the completion?
         1. Due date of each phase or task.

   b. Discuss in detail the human resources needed for the project
      i. How many people, in total, are assigned to work on the project?
      ii. How many people are assigned to work on each major phase of the project?
      iii. For each phase, who does what?

6. **Conclusion**
   a. Provide a short paragraph to express opinions about how the project will be done to conclude the report.
      i. For example: There are many challenges, but they will be successful with great effort.

**SUBMISSION REQUIREMENTS: PART I**:
--) Submit the system analysis report of the generative AI Q&A-Search system.

## 6. PART II: Generative AI Q&A-Search System: High Level & Detailed Design (50 Points)

**IMPORTANT NOTES**:
*--) To **get credit for PART VI**, the student **must take notes of the lecture (in class) on Wednesday 03/24/2024** and **use the notes to design** the Q& –Search system developed with the project.*
*To complete this section, the student **cannot** use any content or materials obtained from the Internet or any external source.*

### 6.1 Q&A Search System: High Level Design

**TO-DO**
- Design (high-level) the semester project Q&A-Search system for the semester project

**SUBMISSION REQUIREMENT: PART II #1**:

- Submit the high-level design of the Q&A-Search System for the semester project

### 6.2 Detailed Design

**TO-DO**
- Design (detailed) the Q&A-Search system for the semester project

**SUBMISSION REQUIREMENT: PART II #2**:

- Submit the detailed design of the Q&A-Search System for the semester project

## 7. PART III: Teamwork Evaluation (10 Points)

**SUBMISSION REQUIREMENTS PART VI**:

Provide the information about your group activities by answering the following questions:
1. What group do you belong to? (Provide the group number)
2. Who are the members of your group?
3. Have the members organized meetings (ONLINE or IN-PERSON) to work on HW 4?
4. If **YES to #3**, which members, including the student himself/herself, showed up in the meeting?
5. If **YES to #3**, do all the members make reasonable efforts to participate actively in the group work?
6. If **NO to #5**, do you have any opinions to share about the group?

## 8. HOWTO Submit

The student must submit all the sections, i.e., submission requirements, in a Microsoft Word document sent to the instructor (Thuan.Nguyen@unt.edu ) as an attachment to a UNT email.

The subject of the email must be:
- "ADTA 5770: Assignment 5 – Submission."

**Due date & time: 11:00 PM – Wednesday 04/09/2025**