MSA 8700: Building Generative AI Business Solutions

Instructors: Péter Molnár and Mark Jack

Term: Spring 2025

Contents

1 Course Details

This hands-on course focusses on developing generative AI applications. It covers large language models, prompt engineering, and fine-tuning frameworks. Students will explore retrieval-augmented generation, document processing, and vector databases. The course also delves into the architecture and deployment of agentic AI applications, image generation, and safeguarding AI systems. Through class activities, students will apply these concepts to real-world scenarios, ensuring a comprehensive understanding of generative AI's business applications.

Day:	Monday
Time (Section 1):	2:30 PM - 5:00 PM
Time: (Section 2):	6:00 PM - 8:30 PM
Location:	Classroom 501, Buckhead Center

1.1 Instructor

Instructors:	Drs. Péter Molnár and Mark Jack
Email:	pmolnar@gsu.edu, mjack6@gsu.edu
Office Phone:	+1.404.413.7713 (Molnár)
Office hours:	TBA & by Appointment
Office Buckhead:	Buckhead Center, 8th floor
Office Downtown:	55 Park, Room #1646

During the term, it is highly recommended that you contact the instructor, in-person or via email. They are available to help you focus your projects, gain access to resources, and answer your questions. Please try to contact them at least once during the term to discuss your project. Your class members are also a good source of help.

1.2 Course Website

Class information will be posted on the iCollege site. There will be links to other websites with course related material.

2 Overview

This course offers an immersive, hands-on experience for students eager to delve into the world of generative AI technologies. This advanced course is intended for individuals who have already mastered the theoretical foundations of deep learning and large language models, as it builds directly on these concepts to explore their practical applications in business settings. Throughout the course, students will engage with a variety of topics essential to the development and deployment of generative AI applications. The journey begins with an exploration of large language models, providing a solid understanding of their capabilities and limitations. Students will then delve into the art of prompt engineering, learning how to craft effective prompts that optimize model performance. This skill is crucial for tailoring AI outputs to specific business needs. As the course progresses, students will investigate

frameworks for prompt evaluation and optimization, ensuring that their AI models are both efficient and effective. Training and fine-tuning large language models is another critical focus, as students learn to adapt pre-existing models to new tasks and datasets. This involves mastering fine-tuning frameworks that enhance model accuracy and relevance. Benchmarking techniques are introduced to equip students with the tools needed to evaluate model performance rigorously. This knowledge is vital for ensuring that AI applications meet the high standards required in business environments. The course also covers retrieval-augmented generation (RAG), a cutting-edge approach that combines retrieval mechanisms with generative models to produce more informed and contextually aware outputs. Document processing for RAG is explored, providing students with the skills to handle and integrate large volumes of data effectively. Vector databases are introduced as a means of storing and retrieving complex data structures, enhancing the scalability and efficiency of AI applications. In the latter part of the course, students will examine the architecture of agentic generative AI applications, focusing on distributed and scalable deployment strategies. This knowledge is crucial for implementing AI solutions that can operate efficiently across various platforms and environments. The course also addresses the ethical considerations of AI deployment, including bias detection, privacy concerns, and federated learning. These topics ensure that students are not only technically proficient but also mindful of the broader implications of their work. The culmination of the course is a comprehensive project presentation, where students will demonstrate their ability to integrate and apply the concepts learned throughout the program. This project serves as a testament to their readiness to tackle real-world challenges in the field of generative AI.

Prerequisite: MSA 8010 - Data Programming for Analytics or IFI 8410 - Introduction to Programming and Predictive Analytics for Business

2.1 Learning objectives

Upon successful completion of this course, you will accomplish the following objectives and outcomes. Students who complete this course will gain “Ready for work” skills, including:

1. Develop a comprehensive understanding of large language models and their applications.
2. Master prompt engineering techniques to optimize AI model outputs.
3. Evaluate and optimize prompts using advanced frameworks.
4. Fine-tune large language models for specific tasks and datasets.
5. Implement benchmarking techniques to assess AI model performance.
6. Utilize retrieval-augmented generation for enhanced AI outputs.
7. Process documents effectively for use in RAG systems.
8. Understand and apply vector databases for efficient data management.
9. Design and deploy scalable generative AI applications.
10. Address ethical considerations, including bias detection and privacy in AI systems.
11. Demonstrate the ability to integrate and apply learned concepts through a final project presentation.

3 Schedule

The course schedule is shown in Table 1. However, the topics and readings may change according to the interests and abilities of the class. Materials may be updated 24 hours prior to class.

Table 1: Course Schedule

Week	Topics	Reading	Assignment	
1	LLM Overview	*AG* 1–3		
2	Prompt Engineering	*AG* 7	Quiz 1	
3	Framework for Prompt Evaluation and Optimization	*KH* 3	Quiz 2	
4	Training and Finetuning of LLMs		Quiz 3	Homework 1
5	LLM Fine-tuning Frameworks, Benchmark Techniques	*AG* 11–12, *KH* 5	Test 1	
6	Retrieval Augmented Generation (RAG)	*AG* 8	Quiz 4	Homework 2
7	Document Processing, Vector Databases	online resources	Quiz 5	
8	Agentic AI Applications	*KH* 4+6	Quiz 6	Homework 3 for RAG
9	AI Agent Frameworks	online resources	Quiz 7	
10	Symbolic and Neuro-symbolic AI	online resources	Test 2	Homework 4
11	Knowledge Graphs	online resources	Quiz 8	
12	Image Generation, Multi-modal Generative Models	online resources	Quiz 9	
13	Safeguarding LLMs, Detecting BIAS, Privacy, Federated Learning	*KH* 10	Quiz 10	
14	GenAI Project Presentations			
				Final Exam

4 Required Resources

4.1 Textbooks

You need access to two textbooks:

AG Jay Alammar, Maarten Grootendorst, *Hands-On Large Language Models: Language Understanding and Generation*, 1st ed., Published by O'Reilly Media, Inc., ISBN-13 978-1098150969.

KH Ken Huang, *Practical Guide for AI Engineers*, Independently published (May 18, 2024), ISBN-13 979-8325962455.

Additional readings will be provided in class.

4.2 Compute Requirements

Programming activities will be performed on the Analytics Research Cluster (ARC) using open-source software and libraries or selected cloud environments.

API Access to Hosted GenAI Models: Limited access to LLMs and other GenAI models will be provided. However, resources may impose wait-times under heavy traffic. You may consider signing up for cloud hosted APIs like OpenAI (`https://platform.openai.com/docs/overview`). Google Colab `https://colab.research.google.com/` or AWS SageMaker Studio Lab (`https://studiolab.sagemaker.aws/`) provide access to GPU supported notebooks and processing environments.

WebEx capable device: You may use any device (laptop, desktop, or tablet) that supports WebEx and other common video conferencing applications. Your device must be capable of transmitting your voice and video to actively participate in the class.

Laptop or desktop computer to complete assignments: ARC provides a web-interface that supports Google Chrome, Firefox and other common web-browsers. A full keyboard and trackpad or mouse are needed to efficiently write code. The browsers on Chrome Books and iPads my not fully support the web-interface.

Print and Scanner: If you cannot take the in-class tests and exams in person, you need access to a printer for your exam forms, and a scanner to digitize your completed forms into a PDF document. The printable forms will be available for download at the time the exam starts. The scanned PDF needs to be uploaded no later than 30 minutes after the end of the exam.

Virtual Private Network (VPN): VPN access is needed to access ARC and other compute resources. Visit `https://technology.gsu.edu/technology-services/cybersecurity/virtual-private-network/` to configure access to the GSU-VPN on your device. (You may seek help from the Technology Service Desk.)

5 Evaluation

Students are evaluated by the deliverables summarized in the table below:

Assignment	Percentage
Quizzes (8×)	10%
Tests (2×)	20%
Final Exam	10%
Homework (4×)	40%
Group Project	20%
Total	100%

Grading Scale:

A+	A	A-	B+	B	B-	C+	C	C-	D	F
97.0%-100%	91.0-96.9	89.5-90.9	87.0-89.4	83.0-86.9	79.5-82.9	77.0-79.4	72.0-76.9	69.5-71.9	60.0-69.4	Below 59.9

6 Homework

Homework assignments in this course are individual tasks designed to enhance student learning and provide opportunities for independent practice. Detailed instructions for completing each assignment, along with submission guidelines, will be made available on the class website. It is essential for students to adhere strictly to these submission instructions, as assignments that fail to comply or are submitted past the deadline will not be considered for grading. This policy ensures fairness and consistency in the evaluation process while emphasizing the importance of personal responsibility in meeting academic expectations.

7 Group Project

Group project assignments are collaborative efforts where teams of up to six members work together to develop an AI application. At the outset of the project, each team is required to submit a group charter that clearly defines the roles and contributions of each member, ensuring accountability and effective collaboration. Teams must use GitHub or the internal GitLab repository to manage their project, with instructors and TAs granted access from the beginning to provide guidance and support. While teams may choose to keep their GitHub repository private, the project serves as an excellent opportunity to showcase skills publicly or as a portfolio piece for job interviews. The AI application developed should include a web interface and be hosted internally, enabling classmates to review and interact with the demo version of the project. This assignment not only fosters teamwork but also helps students build practical skills and create a tangible artifact for future professional use.

8 Tests and Final Exam

Tests and exams will be administered in class using pen and paper. Pre-printed paper forms will be provided. You need to bring a pen or pencil to class. The forms will be scanned and processed for grading.

Students who are not able to take tests and exams in class need to print and scan the completed forms and submit properly scanned documents as PDF files. Instructions on printing, scanning and testing the scanner quality will be provided.

9 Quizzes

Quizzes are administered on-line (via iCollege), during the first **ten** minutes of class. They comprise questions about the topics discussed in the previous class, and topics from the assigned reading.

10 Use of Internet resources and Generative AI

The purpose of assignment is to practice what you learned and verify your understanding of concepts. **The use of Generative AI or any other tools and resources is prohibited during quizzes, tests, and exams.** You are encouraged to utilize Internet resources (like `https://stackoverflow.com/`) and GenAI tools (like ChatGPT, Amazon Q, and GitHub Copilot) for your homework and project assignments. If you make use of these tools indicate in your program code where you found (parts of) the solution or the AI coding tool and prompt that produced the code segment.

11 Student Expectations and Class Policies

Students should plan for 2 - 3 hours of work outside of class each week for each course credit hour. Thus, a 3-credit course averages between 6 and 9 hours of student work outside of the classroom, each week. See GSU site for Academic Success: `https://success.students.gsu.edu/`

11.1 Arbitration

There will be a one-week arbitration period after graded activities are returned. Within that one-week period, you are encouraged to discuss any assumptions and/or misinterpretations that you made on the activity that may have influenced your grade.

11.2 Attendance

If you are unable to attend a class session, it is your responsibility to acquire the class notes, assignments, announcements, etc. from a classmate. The instructor will not give private lectures for those that miss class.

11.3 Submission of Deliverables

Unless specific, prior approval is obtained, no deliverable will be accepted after the specified due date.

If you have a legitimate personal emergency (e.g., health problem) that may impair your ability to submit a deliverable on time, you must take the initiative to contact the instructor before the due date/time (or as soon after your emergency as possible) to communicate the situation. Make-up exams will not be given: However, if a student has a planned absence, he or she may take the exam earlier with the permission of the instructor.

All assignments must be submitted using the designated mechanism that is specified in the assignment (usually via iCollege or ARC). Assignments via email will not be accepted.

11.4 Student Behavior

Behavior in class should be always professional. People must treat each other with dignity and respect for scholarship to thrive. Behaviors that are disruptive to learning will not be tolerated and may be referred to the Office of the Dean of Students for disciplinary action.

11.5 Discrimination and harassment

Discrimination and/or harassment will not be tolerated in the classroom. In most cases, discrimination and/or harassment violates Federal and State laws and/or University Policies and Regulations. Intentional discrimination and/or harassment will be referred to the Affirmative Action Office and dealt with in accordance with the appropriate rules and regulations. Unintentional discrimination and/or harassment is just as damaging to the offended party. But it usually results from people not understanding the impact of their remarks or actions on others, or insensitivity to the feelings of others. We must all strive to work together to create a positive learning environment. This means that everyone should be sensitive to the feelings of others, and tolerant of the remarks and actions of others. If you find the remarks and actions of another individual to be offensive, please bring it to their attention. If you believe those remarks and actions constitute intentional discrimination and/or harassment, please bring it to my attention.

12 Teams of Group Projects

Team Management: Early in the semester, teams will form. If there are problems during the semester, the following methods will be used:

- *Terminating team members:* As in any organization, there may be people in your group who are not willing or able to perform to the level of excellence demanded by the team. The process used to improve team member performance and/or to terminate a team member's membership in the team will involve the following steps:
  - Discuss the poor performance with the individual and the standards he or she is expected to meet. As a team, document the discussion including all members' agreed-upon understanding of the standards of performance and the individual's shortfall from those standards. The document should describe what the individual must do to meet the team's standards and the time frame in which the individual will come up to the standards. This agreement should be signed by all team members, and a copy should be sent to the instructors.
  - If the agreement is not met, the team, including the individual in question, will schedule a meeting with the faculty. The team will bring a copy of the contract to the meeting for the faculty and will discuss the individual's performance with the faculty. The individual will be terminated or given a final chance to improve his or her performance during that meeting and within a given time frame.
  - If the performance does not improve within the time frame, the individual will be terminated from the team.
  - If the individual is terminated, the individual may seek to join another team. Alternatively, he or she must complete all course work in its entirety by himself or herself from that point

forward.

- *Resigning from a team:* A student may resign from a team and switch to a different one. The work that was done while a team member is the property of both the team and the individual so all can use the work product. Faculty will facilitate the placement of the resigning person on a different team.

Teams will be allowed for some activities during the term. Please note that unless the activity is explicitly identified as a "team activity", I expect everyone to perform their own work (your hands on the keyboard). For team activities, you will be allowed to work with partners (of your choosing).

- Initial teams must be established by the second week of classes. Established teams may continue working together on subsequent team activities. Team membership may change during the term if problems arise. However, team members must be designated within one week of the due date for the team activity. Exception: you may withdraw from a team at any time and submit an assignment individually.
- Teams will submit one assignment for all team members. In most cases, each member of the team will get the same score. However, an individual's score may be reduced at the discretion of the instructor.
- Each team assignment must include the following:
  - Tasks completed by each member.
  - Percentage of the total work completed by each member.
- Any individual with a low team contribution will be removed from their team.

13 Official Department and University Policies

1. Prerequisites are strictly enforced. Students failing to complete any of the prerequisites with a grade of "C" or higher will be administratively withdrawn from this course with loss of tuition fees. There are no exceptions, except as granted by the instructor with the approval of the department.
2. Students are expected to attend all classes and group meetings, except when precluded by emergencies, religious holidays, or bona fide extenuating circumstances.
3. Students who, for non-academic reasons beyond their control, are unable to meet the full requirements of the course should notify the instructor, by email, as soon as this is known and prior to the class meeting. Incompletes may be given if a student has ONE AND ONLY ONE outstanding assignment.
4. A "W" grade will be assigned if students withdraw before mid-semester if (and only if) they have maintained a passing grade up to the point of withdrawal. Withdrawals after the mid-semester date will result in a grade of "WF". See the GSU catalog or registrar's office for details.
5. Spirited class participation is encouraged and informed discussion in class is expected. This requires completing readings and assignments before class.
6. All exams and individual assignments are to be completed by the student alone with no help from any other person.
7. Collaboration within groups is encouraged for project work. However, collaboration between project groups will be considered cheating.

8. Copying work from the Internet without a proper reference is considered plagiarism and subject to disciplinary action as delineated in the GSU Student Handbook.
9. Any non-authorized collaboration will be considered cheating, and the student(s) involved will have an Academic Dishonesty charge completed by the instructor and placed on file in the Dean's office and the CIS Department. All instructors regardless of the type of assignment will apply this Academic Dishonesty policy equally to all students. Abstracted from GSU's Student Handbook Student Code of Conduct "Policy on Academic Honesty and Procedures for Resolving Matters of Academic Honesty" `https://codeofconduct.gsu.edu/`.

As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. The University assumes as a basic and minimum standard of conduct in academic matters that students be honest and that they submit for credit only the products of their own efforts. Both the ideals of scholarship and the need for fairness require that all dishonest work be rejected as a basis for academic credit. They also require that students refrain from any and all forms of dishonorable or unethical conduct related to their academic work.

Students are expected to discuss with faculty the expectations regarding course assignments and standards of conduct. Here are some examples and definitions that clarify the standards by which academic honesty and academically honorable conduct are judged at GSU.

13.1 Plagiarism

Plagiarism is presenting another person's work as one's own. Plagiarism includes any paraphrasing or summarizing of the works of another person without acknowledgment, including the submitting of another student's work as one's own. Plagiarism frequently involves a failure to acknowledge in the text, notes, or footnotes the quotation of the paragraphs, sentences, or even a few phrases written or spoken by someone else. The submission of research or completed papers or projects by someone else is plagiarism, as is the unacknowledged use of research sources gathered by someone else when that use is specifically forbidden by the faculty member. Failure to indicate the extent and nature of one's reliance on other sources is also a form of plagiarism. Any work, in whole or part, taken from the Internet or other computer-based resource without properly referencing the source (for example, the URL) is considered plagiarism. A complete reference is required in order that all parties may locate and view the original source. Finally, there may be forms of plagiarism that are unique to an individual discipline or course, examples of which should be provided in advance by the faculty member. The student is responsible for understanding the legitimate use of sources, the appropriate ways of acknowledging academic, scholarly, or creative indebtedness, and the consequences of violating this responsibility.

13.2 Cheating on Examinations

Cheating on examinations involves giving or receiving unauthorized help before, during, or after an examination. Examples of unauthorized help include the use of notes, texts, or "crib sheets" during an examination (unless specifically approved by the faculty member) or sharing information with another student during an examination (unless specifically approved by the faculty member). Other examples include intentionally allowing another student to view one's own examination and collaboration before or after an examination if such collaboration is specifically forbidden by the faculty member. The use of Generative AI tools during quizzes and exams is prohibited.

Unauthorized Collaboration. Submission for academic credit of a work product, or a part thereof, represented as its being one's own effort, which has been developed in substantial collaboration with another

person or source or with a computer-based resource is a violation of academic honesty. It is also a violation of academic honesty knowingly to provide such assistance. Collaborative work specifically authorized by a faculty member is allowed.

13.3 Falsification.

It is a violation of academic honesty to misrepresent material or fabricate information in an academic exercise, assignment or proceeding (e.g., false, or misleading citation of sources, the falsification of the results of experiments or of computer data, false or misleading information in an academic context in order to gain an unfair advantage).

13.4 Multiple Submissions.

It is a violation of academic honesty to submit substantial portions of the same work for credit more than once without the explicit consent of the faculty member(s) to whom the material is submitted for additional credit. In cases in which there is a natural development of research or knowledge in a sequence of courses, use of prior work may be desirable, even required; however, the student is responsible for indicating in writing, as a part of such use, that the current work submitted for credit is cumulative in nature.