CSYE 7374: Special Topics in Computer Systems Engineering [Fall 2025]

Course Information
Course Title: Special Topics in Computer Systems Engineering
Topic: LLMOps: Systems, Infrastructure, and Operations for Large Language Models

Course Number: CSYE 7374
Term and Year: Fall 2025
Credit Hour: 4
Course Format: Traditional

Oakland, CA	Silicon Valley, CA
CRN: 21185 **Location:** Mills Hall 133 **Time:** R 10:35 am - 1:55 pm PT	**CRN:** 21390 **Location:** San Jose 902 **Time:** S 1:30 pm - 4:50 pm PT

Instructor Information
Full Name: Raja Alomari, PhD
Email: r.alomari@northeastern.edu (**Please use messaging functionality in Canvas**).
Office Hours: Wed: 3:00 - 4:00 PM Clickable link (Sept 3, 2025 - Dec 14, 2025).
Contact: **Please use messaging functionality in Canvas.**

Teaching Assistant Information

Oakland, CA	Silicon Valley, CA
TA: TBD **Location:** MS Teams **TA Office Hours:** TBD **Meeting Link: TBD** **Meeting ID:** TBD **Passcode:** TBD	**TA:** TBD **Location:** MS Teams/In person. **TA Office Hours:** TBD **Meeting Link:** TBD **Meeting ID:** TBD **Passcode:** TBD

Course Prerequisites
No official prerequisite. Background in LLMs, MLOps, Cloud Computing is helpful.

Course Description

Large language models Operations (LLMOps) is an emerging discipline at the intersection of machine learning, distributed systems, and cloud infrastructure. This course provides the current operational practices and engineering principles for deployment and lifecycle management of large language models (LLMs). Students will gain the understanding of the full LLM lifecycle including foundational architectures, model fine-tuning, deployment, monitoring, and maintenance in production environments.

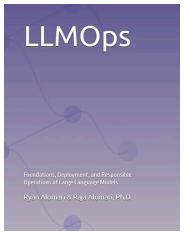
The course covers core topics including LLM architecture and design, differences between LLMOps and traditional MLOps, popular public cloud-based LLM platforms, continuous integration and deployment (CI/CD) pipelines, retrieval-augmented generation (RAG), promptOps, guardrails for safe LLM usage, and monitoring. This course will also cover managing LLMs in private and hybrid clouds.

This course includes multiple hands-on exercises leveraging LLM and ML-related services in AWS, as well as various demos and exercises for different components of LLMOps. In addition, it features a capstone group project that will be mentored and managed through weekly scrums for regular updates. This will serve as an exercise in agile product development, similar to practices in the industry.

Course Learning Outcomes

- Design and manage the full lifecycle of large language models (LLMs), including architecture, fine-tuning, deployment, monitoring, and maintenance in production environments.
- Differentiate LLMOps from traditional MLOps, and apply specialized operational practices tailored for large language models in both public and private cloud settings.
- Implement continuous integration and deployment (CI/CD) pipelines for LLM-based applications, ensuring scalable and reliable model updates.
- Leverage AWS and other cloud services to build, deploy, and maintain LLM solutions, using real-world tools and infrastructure.
- Integrate retrieval-augmented generation (RAG) techniques to enhance LLM performance and contextual relevance in applications.
- Apply prompt engineering (PromptOps) and implement safety guardrails to ensure ethical and secure use of LLMs in production.
- Monitor LLM performance using observability tools, and troubleshoot issues related to drift, degradation, or unintended behaviors in deployed models.

Course Textbooks (Recommended). This course will follow selective flow of the following Textbook



- **Book**:
  - Ryan Alomari and Raja Alomari. LLMOps: Foundations, Deployment, and Responsible Operations of Large Language Models. ISBN: 979-8999628015 (Other formats are available)..
  - The instructor will make a physical copy available in the library.
  - All slides will also be available for all students.

Tools: The following access and tools will be required to perform your assignments. Instructor will provide this access (pending available resources):

- Access to AWS Academy courses:

    - Machine Learning Foundations
    - AWS Learner Lab
    - Machine Learning for Natural Language Processing
- Access to a private cloud environment such as Pextra CloudEnvironment® or Proxmox®, with support for **GPU passthrough** and **virtualized GPUs**, is required. Details will be provided. Instructor will provide free access to Pextra CloudEnvironment® pending available resources. This environment may not have GPU capabilities.

Course Schedule/Topics Covered.

The following is the tentative schedule for this course. Please note that the provided date in each row is for the **start of the week** and not the actual class meeting date.

Order	Week	Topics	HW & Quizzes
1 + 2	Sept 3 Sept 8	Overview & logistics Module 1: Foundation AI Module 2: Foundation LLMs **Hand-on:** Access: AWS Academy AWS Academy Machine Learning Foundations - (SageMaker AI sandbox). AWS Academy Machine Learning for Natural Language Processing - (Module 8). AWS Academy Lab - $50 Limit.	Quiz M1-1 **Option 1:** **Hands-on Lab:** ML*: Lab 3.1 - Amazon SageMaker - Creating and importing data **Option 2:** Assignment 1: Data Collection and Preprocessing for Foundation Model Pre-Training
3	Sept 15 Sept 22	Module 3: Foundation LLMOps **Hands-on:**	Quiz M2-1 **Option 1:**

		Overview relevant AWS services to ML, LLMs, MLOps, and LLMOps.	ML*: Lab 3.2 - Amazon SageMaker - Exploring Data ML*: Lab 3.3 - Amazon SageMaker - Encoding Categorical Data **Option 2:** Assignment 2: Building a Small-Scale Foundation Model from Scratch
	Sept 29	Module 4: LLM Lifecycle Management **Hands-on:** Create a simple foundation Model.	Quiz M3-1 **Option 1:** ML: Lab 3.4 - Amazon SageMaker - Training a model ML*: Lab 3.5 - Amazon SageMaker - Deploying a model ML*: Lab 3.6 - Amazon SageMaker - Generating model performance metrics **Option 2:** Assignment 3: Scaling Up Pre-Training with Distributed Computing

	Oct 6	Module 5: PromptOps and Interaction Management Module 6: Safety, Ethics, and Guardrails **Hands-on:** continue a simple foundation Model.	Quiz M4-1 **Option 1:** ML*: Lab 6 - Amazon Lex - Create a chatbot **Option 2:** Assignment 4: Fine-Tuning a Pre-Trained Foundation Model
	Oct 13	Module 7: Deployment Models and Platforms for LLMs Module 8: Containerization and Orchestration for LLMs **Hands-on:** LLMOps pipeline.	Quiz M5-1 **Option 1:** NLP*: Lab 7.1 - Implementing a multi-lingual solution **Option 2:** Assignment 5: Implementing Retrieval-Augmented Generation (RAG)
	Oct 20	Module 9: Monitoring and Observability of LLM Systems Module 10: Inference Optimization and Deployment Scaling **Hands-on:** LLMOps pipelineMonitoring tools.	Quiz M6-1 **Project Groups and proposals.**

	Oct 27	Module 15: Retrieval-Augmented Generation (RAG) and Vector Databases **Hands-on:** RAG and Vector Databases.	Quiz M7-1 **Scrum:** Project Groups and proposals.
	Nov 3	**Midterm Exam (In person)**	Quiz M8-1 **Scrum:** Project Groups and proposals.
	Nov 10	Module 16: LLM Integration with Enterprise Architectures **Hands-on:** Case study in LLMOps integration..	Quiz M9-1 **Scrum:** Project Groups and proposals. Assignment 6: Model Evaluation and Bias Mitigation
	Nov 17	Module 17: Security and Robustness in LLMOps Module 18: Evaluating LLMs in Production **Hands-on:** Case study in Security Practices in LLMOps integration..	Quiz M10-1 **Scrum:** Project Groups and proposals. Assignment 7: Deployment and Monitoring of LLM System**S**
	Nov 24	Thanksgiving break	

	Dec 1	Module 19: Emerging Trends and Tooling in LLMOps **Project Demos** **Last Day to present any assignment.**	Quiz M12-1 **Scrum:** Project Groups and proposals.
	Dec 8	**Project Demos** **Final Exam (in Person)**	

ML*: Refers to the AWS Academy course: Machine Learning Foundations.

NLP*: Refers to the AWS Academy course: Machine Learning for Natural Language Processing

Scrum: Agile framework for managing and completing complex projects through iterative development, collaboration, and continuous improvement. Book: “Scrum: The Art of Doing Twice the Work in Half the Time” by Jeff Sutherland and J.J. Sutherland

Note: If the lecture falls on a holiday or canceled for any reason, the topic moves directly to the week after.

Note: Quizzes are multiple choice using Canvas (Quiz functionality). Quizzes will be scheduled outside the classroom.

Note: All completed assignments are to be presented in class. One assignment at a time. Deadlines will be provided on Canvas.

Grading Policy:

Category	Weight	Comment
Hands-on Labs (AWS Academy) OR Assignments	25%	Login to AWS course and perform the specific required lab.
Quizzes	10%	Assigned on Canvas.
Midterm	15%	In Person Exam
Final Exam	25%	In Person Exam
Project	25%	Scrums + Final Project Demo.
TOTAL	**100%**	

Grading Scale:

95 - 100% A	89 - 91% B+	79 - 81% C+	Below 72% F
	85 - 88% B	75 - 78% C	
92 - 94% A-	82 - 84% B-	72 - 74% C-	

AWS Academy: Each student will receive access to two courses on AWS Academy. It is similar to Canvas but hosted by AWS. Once you receive the invitation at your Northeastern email address, you will need to enroll. The two courses are listed in the Resources section. Moreover, a controlled lab with a $50 cap will be provided for each student.

Private Cloud Access: The private cloud access will be limited pending available GPU clusters with Private CloudEnvironment. More details will be shared during the class.

Class Recording Policy:

Recording of any kind, including voice, and/or video, and,or photos, is ****strictly prohibited****. This policy complies with applicable laws and respects the privacy and intellectual property rights of all participants.

Attendance/Late Work Policy

Attendance Policy

Students registered in MGEN courses (INFO, CSYE, and DAMG) are allowed **a maximum of 2 absences per course, with 3 or more absences resulting in an automatic 'F' for that course.** Students are expected to inform their instructors of any absences in advance of the class; if a student is sick long-term or experiences a medical issue that prevents class attendance, it is strongly encouraged that they speak with their Academic Advisor (coe-mgen-gradadvising@northeastern.edu) to learn more about the Medical Leave of Absence. Should a student anticipate being unable to attend 3 or more classes, they should discuss their situation with their Academic Advisor to explore other types of leave in accordance with the University's academic and global entry expectations. International students should review the Office of Global Services webpage to understand their visa compliance requirements.

Teaching Assistants (TAs) or Instructional Assistants (IAs) will be present at each class to collect student attendance.

Late Work Policy

***** Students must submit assignments by the deadline in the time zone noted in the syllabus. No late work will be accepted. Each student is responsible for proper planning and submitting on time. *****

End-of-Course Evaluation Surveys

Your feedback regarding your educational experience in this class is particularly important to the College of Engineering. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this course, please take the time to complete the evaluation survey at **https://neu.evaluationkit.com**. Your survey responses are **completely anonymous and confidential**. For courses 6 weeks in length or shorter, surveys will be open one week prior to the end of the courses; for courses greater than 6 weeks in length, surveys will be open for two weeks. An email will be sent to your Northeastern University Mail account notifying you when surveys are available.

Academic Integrity

A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to http://www.northeastern.edu/osccr/academic-integrity-policy/ to access the full academic integrity policy.

MGEN Student Feedback

Students who would like to provide the MGEN unit with anonymous feedback on this particular course, Teaching Assistants, Instructional Assistants, professors, or to provide general feedback regarding their program, may do so using this survey: https://neu.co1.qualtrics.com/jfe/form/SV_cTIAbH7ZRaaw0Ki

University Health and Counseling Services

As a student enrolled in this course, you are fully responsible for assignments, work, and course materials as outlined in this syllabus and in the classroom. Over the course of the semester if you experience any health issues, please contact UHCS.

For more information, visit https://www.northeastern.edu/uhcs.

Student Accommodations

Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.
For more information, visit https://drc.sites.northeastern.edu.

Library Services
The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for education specific resources, visit https://library.northeastern.edu
Network Campus Library Services: Northeastern University Library Global Campus Portals

24/7 Canvas Technical Help
For immediate technical support for Canvas, call 617-373-4357 or email help@northeastern.edu

Canvas Student Resources: https://canvas.northeastern.edu/student-resources/

For assistance with my Northeastern e-mail, and basic technical support:
Visit ITS at https://its.northeastern.edu
Email: help@northeastern.edu
ITS Customer Service Desk: 617-373-4357

Diversity and Inclusion
Northeastern University is committed to equal opportunity, affirmative action, diversity, and social justice while building a climate of inclusion on and beyond campus. In the classroom, members of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration, and an awareness of global perspectives on social justice.

Please visit http://www.northeastern.edu/oidi/ for complete information on Diversity and Inclusion

Title IX
Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty, and staff.

In case of an emergency, please call 911.

*Please visit **https://www.northeastern.edu/ouec** for a complete list of reporting options and resources both on- and off-campus.*

Syllabus Changes: The instructor reserves the right to modify the syllabus, including course content, assignments, grading criteria, and schedule, as deemed necessary to enhance the learning experience or respond to unforeseen circumstances. Any changes will be communicated to students in a timely manner through official channels. It is the responsibility of the students to stay informed about such updates.