

INFO 7375

Prompt Engineering for Generative AI

Course Syllabus

Course Information

Professor: Nik Bear Brown

Email: ni.brown@neu.edu

Office: 505A Dana Hall

Office hours: Through Zoom by Appointment

Note: I am also a master's student at Northeastern. Do not send email to my student email brown.ni@husky.neu.edu as I almost never read that email.

All classes will be held on ground in Boston.

Course website: Canvas

Course Prerequisites

Python programming. A commitment to do independent research.

Course Description

Welcome to the world of Prompt Engineering and Fine-Tuning for Generative AI with Large Language Models (LLMs), where we delve deep into the art and science of crafting prompts that drive LLMs to create captivating and context-aware content. In this comprehensive course, you'll not only master the essential techniques for effective prompt engineering but also gain expertise in the fine-tuning and configuration of LLMs. This dual skill set will empower you to harness the full potential of AI-driven creativity and problem-solving across a wide range of domains.

Course Highlights

- **Prompt Engineering Mastery:** Learn the principles of creating prompts that elicit desired responses from LLMs, whether you're generating text, code, or creative content.
- **Fine-Tuning Expertise:** Explore the intricate process of fine-tuning LLMs, optimizing them for specific tasks, domains, and applications.
- **Real-World Applications:** Apply your skills to real-world scenarios, from content creation and decision support to interactive media and beyond.

- **Ethical Considerations:** Discuss the ethical implications of AI-generated content and responsible AI usage in media production.
- **Hands-On Experience:** Engage in practical exercises, assignments, and projects to reinforce your learning and gain practical experience in prompt engineering and LLM fine-tuning.

By the end of this course, you will not only be proficient in the art of prompt engineering for generative AI but also equipped with the skills to configure and fine-tune LLMs, enabling you to unleash the power of AI-driven creativity and problem-solving across diverse domains. Join us on this transformative journey into the realm of AI-driven creativity and problem-solving.

Learning Objectives

Module 1: Introduction to LLMs and Prompting

- Unveiling Large Language Models (LLMs): Their capabilities, use cases, and historical context.
- Understanding randomness in LLM output and setting the stage for effective prompt engineering.
- Creating Your First Prompts: A hands-on initiation into the world of AI-powered content generation.

Module 2: The Art of Prompt Engineering

- Deciphering the Essence of a Prompt: What is a prompt, and how can it be tailored to your needs?
- Exploring Prompt Patterns: Unraveling the Persona Pattern, Question Refinement Pattern, Cognitive Verifier Pattern, Audience Persona Pattern, and more.
- Applying Prompt Patterns: Crafting prompts for various scenarios, including Few-shot Examples, Chain of Thought Prompting, and Game Play Patterns.

Module 3: Advanced Integration Techniques for LLMs

This module delves into sophisticated methods for augmenting LLMs with vector databases and LangChain. It covers the essentials of vector databases, including their significance in handling high-dimensional data and the process of embedding textual data for enhanced semantic search capabilities. Additionally, the module explores the creation of semantic search applications, focusing on best practices for indexing and querying.

In the segment on LangChain, learners will explore its architecture and utility in building advanced LLM applications, alongside practical setup guidance and development of applications that integrate LLM capabilities with external data sources and APIs. The module also addresses advanced techniques and best practices for LangChain use, culminating in case studies that highlight real-world implementations and solutions to encountered challenges.

Module 4: Fine-Tuning and Configuring LLMs

- **Pre-training Large Language Models:** Unpacking the computational challenges, scaling laws, and domain-specific training.
- **Instruction Fine-Tuning:** Mastering single and multi-task instruction fine-tuning, scaling instruct models, and evaluating model performance.
- **Reinforcement Learning and LLM-Powered Applications:** Aligning models with human values, obtaining feedback, and optimizing for deployment.

Module 5: Agentic AI Systems

- **Foundations of Agentic AI:** Understanding autonomous AI systems that can perceive, reason, and take actions to achieve specific goals.
- **Agent Architectures:** Exploring different frameworks for building AI agents, including ReAct (Reasoning and Acting), ReflexionGPT, and Chain-of-Thought based agents.
- **Tool Use and Function Calling:** Implementing mechanisms for LLMs to interact with external tools, APIs, and data sources.
- **Planning and Reasoning:** Building agents with strategic planning capabilities and sophisticated reasoning mechanisms.
- **Memory Systems for Agents:** Designing short-term and long-term memory solutions to enhance agent capabilities.
- **Multi-Agent Systems:** Creating collaborative environments where multiple specialized agents work together to solve complex problems.
- **Agent Evaluation and Benchmarking:** Methods for testing and measuring agent performance across various dimensions.
- **Ethical Considerations in Agentic AI:** Addressing alignment, safety, and societal implications of autonomous AI systems.

Weekly Schedule

First Quarter (Weeks 1-3)

- Focus on prompt engineering fundamentals and advanced patterns
- Progression from basic prompts to sophisticated techniques like Chain of Thought and ReAct
- Building foundation for understanding how LLMs process and respond to instructions

Second Quarter (Weeks 3-5)

- Overlaps with first quarter, extending prompt techniques
- Introduces theoretical foundations of transformer architecture
- Transitions to vector databases and embedding techniques for knowledge retrieval

Third Quarter (Weeks 6-7)

- Concentrated on LangChain framework
- Focuses on building more complex LLM applications
- Sets up the technical groundwork needed for agentic systems

Final Quarter (Weeks 8-14)

- Extended focus on building agentic systems and final portfolio project
- Covers fine-tuning, RLHF, and deployment strategies
- Emphasizes practical implementation of reinforcement learning in LLM agents
- Culminates in final projects demonstrating agentic AI capabilities

Course Materials

Textbook

- **Title:** "Prompt Engineering for Generative AI" by Nik Bear Brown (Free Online)
- **Publisher:** Abecedarian, LLC
- **Publication Date:** January 2023
- **ISBN:** [ISBN]

Additional Readings

In addition to the textbook, you will also engage with academic papers, AI research reports, and articles specifically related to prompt engineering and fine-tuning techniques in prompt engineering and Generative AI.

By the end of this course, you will not only be proficient in prompt engineering for Generative AI but also equipped with the skills to fine-tune LLMs, enabling you to harness the power of AI-driven creativity and problem-solving across diverse domains. Join us on this transformative journey into the realm of AI-driven content generation.

Course GitHub

The course GitHub (for all lectures, assignments and projects):

https://github.com/nikbearbrown/INFO_7375_Prompt_Engineering_and_Generative_AI

nikbearbrown YouTube channel

Over the course of the semester I'll be making and posting additional data science and machine learning related videos on my YouTube channel.

<https://www.youtube.com/user/nikbearbrown>

The purpose of these videos is to provide additional advanced content as well as supplemental material to enhance coverage of the course topics. Suggestions for additional video topics are always welcome.

Teaching assistants

The Teaching assistants are:

TBA

Programming questions should first go to the TAs. If they can't answer them, the TAs will forward the questions to the Professor.

Learning Assessment

Achievement of learning outcomes will be assessed and graded through:

- Quizzes
- Exams
- Completion of assignments
- Completion of term projects
- Participation (Counts as a 100 point assignment) - TAs will track meaningful contributions to the class and assign a score between 0-100 at the end of finals.
- ATTENDANCE (Counts as a 100 point assignment) - TAs will track whether you are in class. Zoom attendance does not count as attendance.

Reaching out for help

Students can always reach out for help to the Professor, Nik Bear Brown (ni.brown@neu.edu). In an online course, it's important that students reach out early if they encounter any issues.

Grading Policies

A point system is used. Everything you are expected to turn in has points. Points can range from 1 point to 1000 points. Assignments receive a 10% deduction for each day they are late, rounded up. Exams cannot be made up unless arrangements are made before the exam.

I expect to use the following as a rough grading scale at the end of the semester. You should not expect a curve to be applied; but I reserve the right to use one. The curve may go up or down. For example, it is possible for a 95 to be an A-.

Score	Grade
-------	-------

93 – 100	A
----------	---

90 – 92	A-
---------	----

88 – 89	B+
---------	----

83 – 87	B
---------	---

80 – 82	B-
---------	----

78 – 79	C+
---------	----

73 – 77	C
---------	---

70 – 72	C-
---------	----

60 – 69	D
---------	---

<60	F
-----	---

Scores in-between grades (for example, 82.5 or 92.3) will be decided based on the exams.

Note: The score is calculated using the grading rubric and IS NOT the average of the assignments displayed by Canvas.

There will be two major projects due near the end of the semester worth approximately 50% of the total grade. A VFX assignment in Houdini and a game/real-time 3D assignment in either Unreal Engine or Unity 3D.

Canvas

You will submit your assignments via Canvas and GitHub. Click the title of assignment (Canvas -> assignment -> <Title of Assignment>), to go to the submission page. You will know your score on an assignment, project, or test via Canvas. Canvas displays only the raw scores, not normalized or curved grades. A Jupyter notebook file ALONG with either a .DOC or .PDF rendering of that Jupyter notebook file must be submitted with each assignment.

Your name MUST be part of your submission, for example Sanchez_Rick_Assignment_1.zip

Multiple files must be zipped. No .RAR, .bz, .7z or other extensions.

Assignment file names MUST start with student's last name then first name OR the group's name and include the class number and assignment number.

Assignments MUST estimate the percentage of code written by the student and that which came from external sources.

Assignments MUST specify a license at the bottom of each notebook turned in.

All code must adhere to a style guide and state which guide was used.

Due dates

Due dates for assignments are midnight on the due date of the assignment.

Five percent (i.e., 5%) is deducted for each day an assignment is late. Solutions will be posted the following Monday. Assignments will receive NO CREDIT if submitted after the solutions are posted. Any extensions MUST be granted via email and with a specific new due date.

Only ONE extension will be granted per semester.

Course Materials

- Brown, N. B. (2024). INFO 7375 Prompt Engineering and Generative AI. Available for free on GitHub or for purchase on Amazon Kindle for \$0.99.
- Brown, N. B. (2024). How to Speak Bot: Prompt Patterns. Available for free on GitHub or for purchase on Amazon Kindle for \$0.99.

https://github.com/nikbearbrown/INFO_7375_Prompt_Engineering_and_Generative_AI/

AI Policy for Coursework

Use of AI in Assignments

The use of AI tools (such as ChatGPT, Claude, GitHub Copilot, etc.) is permitted in this course, with the following requirements:

- All AI usage must be properly cited
 - Include the name of the AI tool used
 - Specify which portions of your work were AI-assisted
 - Describe how the AI was used (e.g., generating code, editing text, brainstorming ideas)

Demonstration of Understanding

- Students must be able to explain any AI-generated content in their submissions
- Teaching Assistants or the Professor may ask students to walk through and explain any part of their work
- Inability to demonstrate understanding of submitted work may result in grade penalties

Academic Integrity

- Students are responsible for all submitted work, regardless of how it was generated
- AI should be used as a tool to enhance learning, not to bypass it
- Using AI without citation constitutes academic dishonesty

Quality Expectations and Grading

Due to the widespread availability of AI tools, a relative grading component (20% of total points) will be applied to all assignments worth 100 points or more. This component evaluates your work compared to peers, with emphasis on:

- Originality and creativity beyond AI-generated content
- Depth of understanding demonstrated
- Customization and personalization of solutions
- Real-world applicability

Quality Score Breakdown (20 points)

Bottom 25% (5 points)

- Meets basic requirements but lacks depth or real-world relevance

- Basic implementation with minimal customization
- Limited error handling and agent interaction
- Superficial documentation and demonstration

26-50th percentile (10 points)

- Solid implementation with thoughtful agent design
- Functional tools with appropriate integration
- Basic error handling and memory implementation
- Clear documentation and demonstration

51-75th percentile (15 points)

- Strong technical implementation with clear real-world applications
- Sophisticated agent interactions and tool integration
- Comprehensive error handling and memory management
- Professional documentation and compelling demonstration
- Evidence of testing and performance optimization

Top 25% (20 points)

- Exceptional project demonstrating innovation and technical excellence
- Novel application with demonstrable value
- Advanced agent orchestration with sophisticated decision-making
- Unique custom tool that significantly enhances capabilities
- Production-ready implementation with attention to scalability
- Outstanding documentation and presentation

Best Practices for AI Use

Use AI as a learning tool

- Ask for explanations of concepts you don't understand
- Request alternative approaches to problems
- Use AI to review your work and suggest improvements

Maintain your voice and perspective

- Edit and refine AI-generated content to reflect your understanding
- Add your own insights and observations
- Ensure the final work represents your learning and knowledge

Document your process

- Keep a record of your prompts and the AI's responses
- Note any modifications you made to AI-generated content
- Be prepared to discuss your collaboration with AI during evaluation

By following this policy, you can ethically leverage AI tools while ensuring they enhance rather than replace your learning experience in this course.

Participation Policy

Participation in discussions is an important aspect of the class. It is important that both students and instructional staff help foster an environment in which students feel safe asking questions, posing their opinions, and sharing their work for critique. If at any time you feel this environment is being threatened—by other students, the TA, or the professor—speak up and make your concerns heard. If you feel uncomfortable broaching this topic with the professor, you should feel free to voice your concerns to the Dean's office.

Collaboration Policies

Students are strongly encouraged to collaborate through discussing strategies for completing assignments, talking about the readings before class, and studying for the exams. However, all work that you turn in with your name on it must be in your own words or coded in your own style. Directly copied code or text from any other source **MUST** be cited. In any case, you must write up your solutions in your own words. Furthermore, if you did collaborate on any problem, you must clearly list all of the collaborators in your submission. Handing in the same work for more than one course without explicit permission is forbidden.

Feel free to discuss general strategies, but any written work or code should be your own, in your own words/style. If you have collaborated on ideas leading up to the final solution, give each other credit on what you turn in, clearly labeling who contributed what ideas. Individuals should be able to explain the function of every aspect of group-produced work. Not understanding what plagiarism is does not constitute an excuse for committing it. You should familiarize yourself with the University's policies on academic dishonesty at the beginning of the semester. If you have any doubts whatsoever about whether you are breaking the rules – ask!

Any submitted work violating the collaboration policies WILL BE GIVEN A ZERO even if "by mistake." Multiple violations will be sent to OSCCR for disciplinary review.

To reiterate: plagiarism and cheating are strictly forbidden. No excuses, no exceptions. All incidents of plagiarism and cheating will be sent to OSCCR for disciplinary review.

Assignment Late Policy

Assignments are due by 11:59pm on the due date marked on the schedule. It is your responsibility to determine whether or not it is worth spending the extra time on an assignment vs. turning in incomplete work for partial credit without penalty. Any exceptions to this policy (e.g., long-term illness or family emergencies) must be approved by the professor.

Assignments will receive NO CREDIT if submitted after the solutions are posted. Any extensions MUST be granted via email and with a specific new due date.

Only ONE extension will be granted per semester.

Student Resources

Special Accommodations/ADA

In accordance with the Americans with Disabilities Act (ADA 1990), Northeastern University seeks to provide equal access to its programs, services, and activities. If you will need accommodations in this class, please contact the Disability Resource Center (www.northeastern.edu/drc/) as soon as possible to make appropriate arrangements, and please provide the course instructors with any necessary documentation. The University requires that you provide documentation of your disabilities to the DRC so that they may identify what accommodations are required, and arrange with the instructor to provide those on your behalf, as needed.

Academic Integrity

All students must adhere to the university's Academic Integrity Policy, which can be found on the website of the Office of Student Conduct and Conflict Resolution (OSCCR), at <http://www.northeastern.edu/osccr/academicintegrity/index.html>. Please be particularly aware of the policy regarding plagiarism. As you probably know, plagiarism involves representing anyone else's words or ideas as your own. It doesn't matter where you got these ideas—from a book, on the web, from a fellow student, from your mother. It doesn't matter whether you quote the source directly or paraphrase it; if you are not the originator of the words or ideas, you must state clearly and specifically where they came from. Please consult an instructor if you have any confusion or concerns when preparing any of the assignments. You can also consult the guide "Avoiding Plagiarism" on the NU Library Website

at http://www.lib.neu.edu/online_research/help/avoiding_plagiarism/. If an academic integrity concern arises, one of the instructors will speak with you about it; if the discussion does not resolve the concern, we will refer the matter to OSCCR.

Writing Center

The Northeastern University Writing Center, housed in the Department of English within the College of Social Sciences and Humanities, is open to any member of the Northeastern community and exists to help any level writer, from any academic discipline, become a better writer. You can book face-to-face, online, or same day appointments in two locations: 412 Holmes Hall and 136 Snell Library (behind Argo Tea). For more information or to book an appointment, please visit <http://www.northeastern.edu/writingcenter/>.