COMS4995W31 Applied Machine Learning (Spring 26)

Course Description

Applied Machine Learning is not just about training models on clean datasets. In practice, ML systems operate inside complex, messy environments shaped by delayed feedback, data drift, and business constraints.

This course is designed to bridge the gap between algorithms and production. We begin with the "**Classic Stack**" (Linear Models, Trees) and move to the "**Modern GenAI Stack**" (LLMs, RAG, Agents). Unlike theoretical courses, we emphasize problem framing, system design, and failure analysis. We treat models not as mathematical artifacts, but as powerful components to be fine-tuned, chained, and deployed to survive real-world traffic.

Note on Deep Learning & Math**: This course prioritizes Deep Learning and LLMs as the core of modern ML engineering. We cover mathematical foundations specifically to ensure a "white-box" understanding necessary for debugging and optimization, focusing on **applied engineering logic rather than theoretical derivation.

***Note from the Instructor**: My goal for this course is simple: **to bridge the gap between academic theory and the reality of shipping AI at scale**. As a practitioner leading various cutting-edge projects, I have designed this curriculum to simulate the rigorous standards of a top-tier tech environment.*

*You will find some policies in this syllabus strictly enforced (e.g., code quality, system design constraints). **Please understand that this rigor is not meant to penalize you, but to prepare you.** I want you to make mistakes here - in a safe, supportive classroom - so you do **not** make them in your technical interviews. My TA team and I are fully committed to your success and are here to help you land your dream offer in this competitive market*

Learning Objectives

By the end of this course, students will be able to:

- **[Problem Formulation]** Translate ambiguous business requirements into well-defined ML tasks.
- **[System Reliability & Failure Analysis]** Understand why models fail in production (e.g., Data Leakage, Feedback Loops, Reward Hacking) and how to design monitoring systems to catch them.
- **[Modern GenAI Stack]** Implement RAG systems, perform SFT, and design LLM Agents.
- **[Classic ML Proficiency]** Master feature engineering, pipeline construction, and error analysis using scikit-learn.
- **[Engineering Rigor]** Utilize AI coding assistants effectively to ship production-ready code

Prerequisites

- **Programming:** Proficiency in Python 3. Students must be comfortable with data manipulation (`pandas`, `numpy`, `pytorch`) and debugging.
- **Tooling:** Familiarity with AI coding assistants is **strongly recommended**, as they will be used extensively for assignments.
- **Mathematics:**
    - Linear Algebra: Vector/matrix operations, dot products, dimensionality reduction concepts.
    - Statistics: Basic probability, distributions, and hypothesis testing concepts.
    - *Note: While we won't derive every theorem, a mathematical intuition is necessary to understand why models fail.*
- **Mindset:** A willingness to adapt. The ML field changes monthly; course materials may update in real-time to reflect new SOTA (State-of-the-Art) developments

Course Logistics

- Classroom: Fayerweather 310 (Morningside Campus)
- Schedule: Monday, 4:10 pm - 6:40 pm (Spring 2026)
- Credits: 3.0

Instructor

- Dr. Spencer W. Luo
- Email: swl2145@columbia.edu

Teaching Assistants

- Case Hallowell Schemmer (chs2164@columbia.edu)
- Grace Yoon (gy2354@columbia.edu)
- Zoga Duka (zd2377@columbia.edu)
- Prajwal Raghunath (pr2789@columbia.edu)

TA Office Hours: TBD, updated on Ed
Optional Recitation: TA-led sessions for coding and math refreshers

Recommended Textbooks

These books focus on engineering implementation and system design, aligning with the course's applied philosophy:

- **Designing Machine Learning Systems** by Chip Huyen
    - *Focus: Production systems, data pipelines, monitoring, and failure analysis.*
- **Build a Large Language Model (From Scratch)** by Sebastian Raschka.
    - *Focus: A "white-box" coding approach to understanding Transformers and LLM*

Course Schedule

Class	Lectures	Assignment
1	Introduction to Applied Machine Learning	
2	Data Preparation and Feature Engineering	
3	Linear Models **[Industry Case Study]** *Why Linear Regression is still widely used in Ads (Latency constraints)*	Assignment 1 out
4	Model Evaluation & Bias-Variance Focus: Beyond Accuracy (Precision, Recall, AUC)	
5	Tree-based Models & Ensembles **[Industry Case Study]** *Stripe - Why Trees outperform Neural Nets on tabular data*	
6	💪 Midterm Exam - **03/02/2026**	
7	Neural Networks Fundamentals	Assignment 2 out
8	Transformers in Practice **[Industry Case Study]** *The Inference Bottleneck - Balancing Model Size vs. Low Latency (Distillation & Quantization)*	
9	Pre-training & Supervised Fine-tuning	
10	RAG & Vector Databases	Assignment 3 out
11	LLM Agents & Tool Use **[Industry Case Study]** *Autonomous Loop Failures – Why Agents get stuck and how to add guardrails*	
12	Course Review	
13	💪 Final Exam - **04/27/2026**	

Assessments

Philosophy This course simulates an industry environment. We embrace modern AI tools for implementation, while prioritizing strict evaluation of your fundamental understanding through exams.

Course Activities

- **3 Homework Assignments (60%)**
  - Students will work on concrete ML problems covering feature engineering, modeling, and evaluation.
  - **Deliverables**: For each assignment, you must submit a code repository and a brief summary of your results and findings.
  - **AI Policy**: Use of AI tools is **fully permitted** for both code and documentation. You are responsible for ensuring the final output is correct and runnable.
  - **Code:** Must be easily accessible via a runnable link (e.g., Google Colab, GitHub). **Screenshots of code are not accepted and will receive a grade of 0.**
  - **Reports:** Summary of results should include visualizations and written explanations as outlined in each assignment.
- **1 Mid-term Exam (20%)**
  - In-class, closed-book. Covers fundamental concepts from the first half of the course.
- **1 Final Exam (20%)**
  - In-class, closed-book. Covers advanced concepts and holistic understanding of the material

Scope of Assessments

Students are responsible for all material associated with this course. Any content presented, assigned, discussed, demonstrated, or referenced as part of the course is considered fair game for exams. This includes, but is not limited to: lectures (in-person and Zoom), slides, whiteboard discussions, homework assignments, recommended readings, code examples, case studies, guest content, and materials posted on Ed.

That said, **we are not here to trick you.** While the scope is comprehensive, our exams prioritize deep engineering intuition and system design thinking over trivial memorization. If you engage actively with the lectures and assignments, you will be well-prepared. Think of the exams not as hurdles, but as opportunities to demonstrate your readiness for interviews tomorrow.

Grading Scale & Curve

Final letter grades will be determined based on your performance relative to the class distribution. We do not use fixed percentage buckets. The instructor reserves the right to apply an overall curve (upward or downward) based on the collective performance of the class. Scores will be rounded according to standard rounding rules before applying the curve

Course Policies

Zoom Support

Lectures will be streamed via Zoom and recorded on a **"best effort"** basis. Please be aware that technical failures (e.g., internet outages, audio issues) **may result in the loss of both the live stream and the recording**

- **Contingency:** In the event of a technical failure, **students are responsible for obtaining notes from peers**. The instructor will not repeat the lecture
- **Recommendation:** We strongly recommend attending in-person to guarantee access to the material

Attendance

Attendance is **optional and not graded**. We treat you as adults; you do not need to email the instructor to report or explain absences.

Late Work

Late submissions will be accepted up to **24 hours** after the original deadline with a **10% penalty**. Work submitted more than 24 hours past the deadline will **not be accepted** (graded as 0).

Exceptions: We understand that life happens. If you have a busy interview schedule, health issue, or other emergency, please submit a request via the Google Form (to be announced in Ed) before the deadline. **Extension requests sent via email will not be processed.**

Re-grading

Regrade requests must be submitted in writing to the TAs **within one week** of the grade release. Note that requesting a regrade triggers a full review of your assignment, and your score may **increase, decrease, or remain the same**.

AI Policy

Use of AI tools (ChatGPT, Claude, Gemini, etc.) is **permitted** for all assignments. You are expected to verify AI-generated outputs. Code that crashes or logic that fails will be graded accordingly, regardless of its origin.

While AI tools are permitted, **plagiarism of other students' work is strictly prohibited**. You are personally responsible for the accuracy and logic of your submissions. Submitting code you cannot explain or debug during a review will be considered a violation of the honor code.

Students with Disabilities

If you have been certified by Disability Services (ODS) to receive accommodations, please contact **Robert Kramer** (Associate Director of DSI) at rk3281@columbia.edu to arrange strictly confidential logistics (e.g., separate testing rooms). Once your arrangements are confirmed with DSI, please simply notify the TAs so we can update the exam timer settings in our system.

Communication

All class discussions, technical questions, and clarifications should be handled on **Ed Discussion** (**link**). This ensures streamlined support for everyone. Please reserve email solely for sensitive personal matters not covered by the extension form.

Final Thoughts

This course moves fast and covers a lot of ground. It will be challenging, but it will also be rewarding. Remember that you are not alone in this journey - the teaching staff and I are your resources. We value communication and transparency, so please utilize Ed Discussion and Office Hours. Let’s build something amazing together this semester.