ML ENGINEERING CAPSTONE
Syllabus: CSC490 Fall 2025

1. Instructors.

- Denys Linkov
  Email: csc490prof@cs.toronto.edu
  Office Hours: Email to Book

2. Lectures. This course has one three hour section

- CSC412H1-S-LEC5101
- Lectures - DSCIL (2nd floor Gerstien) - Weds 6-9pm

3. Teaching Assistants.

- Yuchi (Allan) Zhao

4. Course webpages. The course webpage contains all course information, additional readings, assignments, announcements, office hours, etc. You are expected to check the following sites regularly!

- diophontine.github.io/csc490/
- q.utoronto.ca
- https://piazza.com/utoronto.ca/fall2025/csc490

5. Course Evaluation.

- 6 assignments: 50%, lowest mark dropped. All assignments are due at 11pm ET on the due date.
  - A1 - 10% Due Sept 17
  - A2 - 10% Due Oct 1
  - A3 - 10% Due Oct 15
  - A4 - 10% Due Nov 5
  - A5 - 10% Due Nov 12
  - A6 - 10% Due Nov 26
- Final Presentation 50%
  - In person + Github submission Dec 2 (Makeup day)

6. Course Outline. This course covers a variety of topics on ML problems, training and inference.

1. The Gen AI Landscape and AI product management
2. Intro to docker, kuberenetes, cloud, terraform and architecture diagrams
3. Evaluating ML products
4. Prompting and Constrained decoding
5. Model serving deep dive 1 - Quantization and CUDA
6. Model serving deep dive 2 - Speculative decoding & KV caching
7. Model training - LLM Pre-training, SFT
8. Model training - LLM Reinforcement learning and Prompting, PPO, GRPO, RLHF
9. Scalability of ML Systems
10. Data Engineering
11. Feature stores and Evaluation metrics
12. Search and Recommender systems
13. Model serving deep dive 3 - LORA and Model Serving

7. Prerequisites.

- 1.5 credits of 300+ level CSC courses.

8. Textbooks. There is no required course textbook. Helpful material will be posted throughout the semester.

9. Assignments. There will be 6 assignments in this course. Assignments will be released on Quercus and submission instructions will be provided with each assignment. These assignments will be related to the course topics and will help structure your project.

9.1. *Assignment structure.* Each assignment will have a combination of pull-requests and artifacts that are submitted. These may include code samples, diagrams, charts, plots, written contributions. These are meant to simulate building and researching a new idea within a team environment. Assignments will be graded by the instructor and TA based on the artifacts provided.

9.2. *Collaboration policy on the assignments.* Assignments will be completed in your project teams and all team members will recieve the same grade on the assignment and project.

10. Exams. There is no exam for this course.

11. Project. The final project will help you build the skills for research or industry based work. You may choose to work on a research project or build a ML focused startup.

11.1. *Objectives.* The goal of your project is to apply course material to solve a problem you care about. The problems are quite open ended up we encourage you to pick something you are motivated to go in depth into. Most of your grade will come from the code you write as part of the final project.

11.2. *Collaboration policy.* You will be working in a team environment with 3-4 students (4 preferable). All students will recieve the same grade for the project unless there are exceptional circumstances.

11.3. *Evaluation.* Evaluation will be based on an in class presentation (20%) and submitted artifacts review (30%)

12. Late policy. Ten percent of the value will be deducted for each late day (up to 3 days, then submission is blocked). No credit will be given for assignments submitted after 3 days.

13. AI policy. You are encouraged to use Gen AI within this course representative of workplace AI usage, and will be evaluated with the corresponding productivity gains in mind.

14. Absence declaration. Students who are absent from academic participation for any reason (e.g., COVID, cold, flu and other illness or injury, family situation) and who require consideration for missed academic work have been asked to record their absence through the ACORN online absence declaration. The absence declaration is considered sufficient documentation to indicate an absence and no additional information or documentation should be required when seeking consideration from an instructor. Students should also advise their instructor of their absence. Instructors will not be automatically alerted when a student declares an absence. It is student's responsibility to let instructors know that they have used the Absence Declaration so that you can discuss any needed consideration, where appropriate.

15. Grading concerns. Any requests to have graded work re-evaluated must be made within one week of the date the grade is released. Re-evaluation may result in a decrease in the grade.

16. Computing. In the assignments and project, you will need to write your own programs, debug them, and use them to conduct various experiments, plot curves, etc. You may use any programming language, but `Python` is preferable. You will also be able to select computing environments to build your projects within.

Students are recommended to use github and join our class organization `https://github.com/UofT-CSC490-F2025` to submit assignments and real-world best practices.

17. Accommodation for Disability Policy. Please send your documented accessibility requests directly to the instructor, at least a week before the due date of each evaluation item. Extensions may be granted, and the duration will be determined based on the letter from the Accessibility Services at the University of Toronto.