Seminar on Artificial General Intelligence (CSC 209/409) - Syllabus
Spring 2025

Instructor & Lecturer: Prof. Christopher Kanan

2 Credit Hours

Catalog Description: Many companies are working towards creating artificial general intelligences (AGIs) and artificial super intelligences (ASIs), i.e., systems capable of accomplishing any human intellectual task. A growing number of companies and AI researchers believe AGI will be achieved within the next decade. While large language models (LLMs) do not yet have these capabilities, they exhibit “sparks” of AGI. In this course we review the concept of AGI/ASI and how we might assess a system to understand if it has achieved that capability. We will study the gaps in state-of-the-art systems and what would be needed for a system to be considered an AGI. We will also study AI ethics and safety, the socioeconomic implications of AGI, the likelihood of an intelligence explosion (i.e., the singularity), and existential risk due to AGI.

Course Frequency: Offered in Spring

Prerequisites: Students should have a good understanding of how large language models are created, especially causal LLMs (e.g., GPT-like models). Students should have at least a high-level understanding of neural networks. Students should have at least a high-level understanding of neural networks (e.g., loss functions, backpropagation, activation functions, transformers, etc.) and machine learning basics (e.g., evaluation, train/test splits, etc.). These will be briefly reviewed early in the semester as a refresher.

Class Location & Time: Mondays*, 2:00 PM - 3:15 PM, Hylan Building Room 203
*Note that class will also be held on Friday January 24 (UR “Monday”)

No textbook is needed for this course, but there will be readings.

Slack Link: https://join.slack.com/t/seminaronagis-jnz6964/shared_invite/zt-2ybue5sat-VeUli7D3bi1nu4Zj38_hJg

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

1. **Understand the Foundations of AGI:** Define AGI and ASI, identify key differences from current AI systems, and evaluate the criteria for achieving general intelligence.
2. **Evaluate the Path to AGI:** Analyze current AI capabilities, identify gaps and challenges, and assess the feasibility and timeline for AGI development.
3. **Assess Risks and Implications of AGI:** Critically examine the ethical, safety, societal, and existential implications of AGI and its potential impact on humanity.

Instructor Contact:

Name: Prof. Christopher Kanan
Office Hours: TBD Wegmans 3017
Email Address: ckanan@cs.rochester.edu

Illness Policy: All lecture slides will be posted online. If you are feeling unwell, you are encouraged to not attend class or at the very least wear a well fitted mask.

Teaching Assistants & Graders:

Role	Name	Office Hours	Email
TA	Yousra Awad	TBD	yawad2@u.rochester.edu

Evaluation and Grading: The final course grade will be weighted as follows:

Exams: 40%
Project: 60%

We will follow standard grading guidelines to assign the percentage into a letter grade. The professor may choose to "curve" the class by giving all students the same number of additional points.

Capstone Project: You are required to complete a project related to the topics covered by this course. This could be a very well researched perspective / position paper, an analysis of existing systems to assess them for properties related to AGI, or an AI research project that is aligned with the topics covered in this course. If your project involves programming, you may use the programming language and toolboxes of your choice. Your write-up should be rigorous. Run your early ideas by Prof. Kanan and other staff via email or in person. Teams may be permitted with permission, but only for projects involving creating or assessing AI systems. The schedule for the project is as follows:

1. **Project Proposal**: The project proposal should clearly state what you plan to do. It should be at least 2 pages long (not including references, figures, or tables). It should contain a list of three to six milestones and deadlines. You should list what software you will be using or will build upon. Describe the datasets you will use and how will you know if the project is successful. Describe the related work. The proposal should be a well organized document in continuous English, and it should not be merely an outline. You should be able to reuse much of the text for the final report. It should be submitted as a PDF (under 10MB).
2. **Revised Project Proposal (optional)**: The revised proposal is an opportunity to improve your grade if you fail to do the project proposal effectively. You may submit a revised proposal that takes into account the comments received by the instructor and/or TA. The new grade will replace the original score, but the maximum score for the revised proposal is 80%.
3. **Project Report**: The project report will describe the project, i.e., what you did and the result. *It should be at least 4 pages long (not including references, figures, or tables)*. The report should be formatted in NeurIPS format (exceptions will be granted if you are submitting to another conference or journal). It should be submitted as a PDF (under 10MB).

Policy on Late Work: No credit will be given for the project report if it is turned in late. For the project proposal, a late submission will be treated as submitting a revised project proposal capping the score at 80%. For homework assignments, full points will be awarded only if the assignment is turned in at most one day late. Late homework assignments will be accepted up to 7 days late with a 20% penalty imposed, meaning the highest possible score will be 80% for any assignment that is 2-7 days late. No credit will be given for assignments turned more than 7 days late. An exception to this policy is that no late work will be accepted after the final assignment's deadline.

Academic Honesty and Integrity: All assignments and activities associated with this course must be performed in accordance with the University of Rochester's Academic Honesty Policy. You are expected to read, understand, and follow the policy. Additionally:

- In general, homework is to be completed independently. However, you are encouraged to study together and to discuss information and concepts covered in lecture and the sections with other students. You can give "consulting" help to or receive "consulting" help from such students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of the work done by someone else. Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.
- Posting homework and project solutions to public repositories on sites like GitHub is a violation of the College's Academic Honesty Policy, Section V.B.2 "Giving Unauthorized Aid."
- During examinations, you must do your own work. Talking or discussion is not permitted during the examinations, nor may you compare papers, copy from others, or collaborate in any way. Any collaborative behavior during the examinations will result in failure of the exam, and may lead to failure of the course and University disciplinary action.

Prior Course Materials: Unauthorized use of course materials from previous semesters (e.g., material you have received from others), is strictly prohibited.

New Course Materials: Course materials (slides, lectures, assignments, etc.) may not be re-distributed or posted elsewhere online. Redistribution of copyright protected material outside this course may be prohibited by law.

Notes on Plagiarism: Plagiarism is a serious offense and is in violation of university policy.

- If you are unsure of what constitutes plagiarism in written documents, a good description can be found here: https://rochester.edu/college/gradstudies/assets/pdf/Plagiarism_Misconduct.pdf
- Plagiarism does not just occur in written documents; it also occurs in code. Many of the algorithms we will code and problems we will solve have been solved by others who have posted code (in various programming languages) online. It is unacceptable (and it is considered plagiarism) to copy code developed by others and submit it as your own. (This includes code that is written by your fellow students!) Even making minor changes, such as changing variable names, function names, formatting, etc., is not enough to allow you to claim your submission as your own because the underlying structure of the code remains unchanged. You may also be in violation if you excessively rely on AI "Co-Pilot" systems to assist you with writing your code or using generative AI to do your assignments.
- If you do consult any online sources of code, you must properly attribute the corresponding sections in your code to their original source, as you would add quotations, footnotes, or references in a written document. The consequences of plagiarism, whether in code or in written documents, are at the discretion of the instructor, and can be as severe as automatic failure of the course.

Academic Accommodations: We are committed to providing reasonable accommodations to students with disabilities. Please see the professor about your required accommodations as early as possible in the term. The University of Rochester respects and welcomes students of all backgrounds and abilities. In the event you encounter any barrier(s) to full participation in this course due to the impact of a disability, please contact the Office of Disability Resources: disability@rochester.edu; (585) 276-5075; Taylor Hall; www.rochester.edu/college/disability.

Course Schedule: The following schedule lists dates for class topics. *The content in this schedule is tentative and subject to change.* It is your responsibility to attend class and to remain informed of any changes that may be announced.

Week	Date	Assignments	Class / Discussion Topics	Presenter
1	1/24 (Fri.)		**Course Introduction:** Overview & Logistics	Kanan
2	1/27		**Primer on Modern AI:** Deep Learning, LLMs, & Generative AI	Kanan
3	2/3		**Why Are Human's General Intelligences?** Kinds of Intelligence in Humans	Kanan
4	2/10		**Criteria for AGI:** How Do We Assess AGI Capabilities?	Kanan
5	2/17		**Capabilities and Gaps**: Current AI systems vs. AGI – limitations and challenges	Kanan
6	2/24		**AI Ethics and Alignment**: Value alignment and societal norms	Kanan
7	3/3	Capstone Proposals Due	**Exam 1:** Covers Weeks 1-6 Topics	N/A
8	3/10		**No Class:** Spring Break	N/A
9	3/17		**Socioeconomic Impacts of AGI**: Labor, wealth redistribution, inflection points, and governance	Kanan
10	3/24		**The Intelligence Explosion & Existential Risk**: Likelihood of a singularity and AGI in Warfare	Kanan
11	3/31		**AI Philosophy & Consciousness**: Computational theories and moral patienthood of AI	Kanan
12	4/7		TBD - Used for Overflow	Kanan
13	4/14		TBD - Used for Overflow	Kanan
14	4/21		**Closing Discussion:** Priorities & Open Questions	Kanan
15	4/28	**Last In Person Meeting**	**Exam 2:** Covers Weeks 9-14 Topics	N/A
N/A	5/8		Capstone Project Report Due	