CSE199 DAIS Homework and Group Presentations Fall 2026

AI Zoom: Boom or Doom?

No shortage of rhymes here—the title could be "AI Vroom: Bloom or Gloom?" or "AI Spume: Loom or Tomb?" This assignment calls for serious contemplation of challenging, yes horrid, scenarios. The goal of the **written homework** is to convey how they have moved from standard fare of fantasy sci-fi apocalypse to concrete particulars already impacting our world. The highlighted sources for this assignment break dangers into several areas and issues within each area, such that individual group members can choose and write on separate areas. The **third-week presentations** will choose from the issue areas and outline possible solutions. The most effective presentations will also include (i) evaluation of obstacles to the solutions and/or (ii) what we can do in the department and university on the whole, from undergrad to faculty levels, to prepare for them.

Two recent flumes of impact have been *AI-guided cyberattacks* represented by July's *HuggingFace* incident, and *AI's conquests in research-level mathematics and CS theory.* The first few lectures in the unit will discuss both. The latter may appear to matter less in the grand global picture, but (i) your instructor has ongoing direct experience with it, and (ii) mathematics has given rise to pervasive technology from atomic bombs to mass-market cryptography.

A key concept is **alignment**. This means the degree to which AI developments advance human welfare. The technical image is the angle between two vectors. The more the vectors point in the same direction, the more positive the cosine of the angle. But if they point away from each other, the cosine is negative and the effects are detrimental. This is a technical concept "under the hood" of training AI models themselves, but the idea matters at top level. We want AI outcomes to align well with—or at least not work against—as many "human value vectors" as possible. To see how complex this becomes, we need only think of mathematics: Advancing human knowledge by any means is innately a boon, but the "AI gold rush" has happened faster than the mathematical community has been able to adapt. A level-headed statement of health for our future with AI is, *Can we solve the alignment problem?*

1 Sources and Issue Categories

The **one required source** is an essay dated August 2026 by Dr. Xiaoyu He titled "Existential Risk From AI: An Exposition For Mathematicians." The essay is not just for mathematicians, but the author, in the Mathematics Department at Georgia Tech, is appealing first and foremost to his own professional community. The direct link is https://alkjash.github.io/ai-risk/, and a PDF mirror is in the course unit folder at https://cse.buffalo.edu/ regan/cse199/XiaoyuHeAIRiskAug2026.pdf. The essay's main virtue is its outline, which lists 20 statements to debate in five issue areas:

1. Immediate, Direct AI Takeover Scenarios.
2. Killer Tech Unlocked By AI (as if by magic).
3. AI Bootstraps Itself Into The Singularity.
4. Alignment Fails Intrinsically.
5. Alignment Fails Because of Human Failings.

One of the essay's major points is that multiple combinations of just one or a few of these can spell doom (the two lone math formulas given mainly serve this point). You, individually, need only be responsible for one of these five areas. Your group of 3-or-4 can take 3-or-4 of these areas, or *perhaps better*, just two of the five areas and divide up the associated statements in each of them.

2 Written Homework, due Week 2 Wed., 11:59pm

For the **individual written homework**, taking one area, it suffices to give **two examples**—apart from any mentioned in the essay itself—that illustrate statements in the area, and/or could bring them about. The examples can be drawn from other sources, including any in the list given below. A recommended format for a 1–2 page essay is to have one paragraph introducing the area, then two paragraphs giving and discussing your examples, and a conclusion giving your own estimation of the danger. (Did your stance change while doing this exercise?)

The sources must be *published humanly written entities*, not AI chats. Note, however, the very bottom of Dr. Xu's essay: "Written with the help of Claude Fable 5 and OpenAI ChatGPT 5.6 Sol." Your written submission should include the sources for your examples as end citations.

Grading: Our of **3 pts.**, there is **1 pt.** for participation. Then **0.5** for choosing an area and describing it further, **0.5** for each of the two issue statements and their examples, and **0.5** for the conclusion with your stance.

3 Recitation Presentations, Thu.-or-Fri. of Week 3

The **group presentations** have a *required extra element*: to propose ways by which the danger from the chosen issues may be averted. The solution-finding is intended to come from collective group discussion. This envisions a presentation structure in which each member first speaks individually for 60-odd seconds about their part of the group's chosen issues, then the group tag-teams possible solutions in the final 2 or so minutes.

It suffices for each speaker to mention **one** of his/her two examples—this allows for possible overlap within the group, too. The slide deck should have a slide that collects all the sources for examples and also at least one human source for potential solutions. If you can come up with ways that beginning CS students can possibly contribute—besides not cheating with AI—so much the better. But this is not required, and this is buffeted by two factors:

- Your faculty elders don't fully know how to contribute toward solutions either.
- Your group is **welcome to ask an AI about a possible solution** to an issue **that you collectively specify**.

The latter is *optional*. The groundrules if you do it are (i) your group must still have the human source for one potential solution as specified above, (ii) you must show the prompt(s) you used, and (iii) your group must include a "15-second critique" of what the AI says.

Grading: The **3 pts.** are divided into the customary global **1 pt.** for participation, then **1 pt.** for individual role in the presentation, and **1 pt.** for each member based on the quality of the collaborative effort, which includes brainstorming and evaluating potential solution(s) in the final presentation segment.

It bears repeating that in both branches of this assignment, you are primarily graded for *engagement*, the effort to digest and summarize unfamiliar information, presentation experience, and development of critical thinking. You are not being graded to come up with correct answers—this assignment in particular doesn't know them, and earnest deviation may be needed more than "the company line." Above all, the parameters aim at having fun—even amid this most ghastly of topics.

4 Other Sources

This list may be expanded as the term progresses—indeed, the first two appeared after this term began. Not every source has examples of issues—some are just commentary. **Of course, you are welcome to seek and use other relevant published sources.** The sources listed here are a guide.

- Bill Gates, "The turbulent AI era is here. The choices we make now are critical." *Gates Notes* weblog, 8/26/2026. Direct link: https://www.gatesnotes.com/a-turbulent-ai-era-and-critical-choices-to-make. You are welcome, however, to read instead the "Cliff Notes" version at the ISOC Live Substack. There is also an interview of Gates about this by the MIT Technology Review.

  Gates highlights three dangers: the effect on jobs, empowerment of bad actors (a-la areas 2 and 5 by Dr. He), and the detriment of human relationships and capabilities. My own earlier units in this course focused on the last—and you may add it as a permissible "area 6" to the five from the required essay.

- Tyler Cowen, "Should You Believe the Dire Warnings About 'AI Takeover'?" *The Free Press*, 8/30/2026. (Original link behind paywall, but mirrored here, and there is a related video. A commenter on X ascribed to Cowen the stance that AI safety creates as many opportunities for CS careers as AI takes away, but I do not find that specifically in this article. Disclaimer: Cowen is a personal friend from childhood.)

- James Dargan, "AI Safety Expert Roman Yampolskiy Believes AI Has a 99.9% Chance of Wiping Out Humanity," *AI Insider*, 7/6/2026. Free link. (Yampolskiy is often credited with coining the term "AI safety." He obtained his PhD at UB in 2008 under Professor Venu Govindaraju.)

- Andrew Critch and Jacob Tsimerman, "A Taxonomy of Omnicidal Futures Involving Artificial Intelligence," ArXiv https://arxiv.org/abs/2507.09369v1, 7/15/2025. (Much heavier going than Dr. He's essay. Tsimerman won one of this year's Fields Medals, then announced he is leaving the University of Toronto's mathematics department to work on AI safety.)

- Liv Boeree,"The Dark Side of Competition in AI," TED talk posted 11/9/2023, direct link. Also see this recent interview. (I already showed parts of this video in my 2024 and 2025 units; it ties in to your "Prisoner's Dilemma" activity in week 1.)

- Richard Lipton and Kenneth Regan, "Human Extinction?" Weblog *Gödel's Lost Letter and P=NP*, 6/8/2023. https://rjlipton.com/2023/06/08/human-extinction/ (Well, I've been jolted off my own stance expressed there. You may ignore the technical second section.)

5 Freedom Note

This assignment description lays out structures and requirements for the sake of definiteness—and to make a rubric for grading that promotes even-handed treatment of students from various backgrounds and experiences. If, however, you happen to have some particular experience that is fully relevant, then you may deviate from strict adherence. For example, you may focus more deeply on just one issue, with just one example made all the richer by your experience. The guidelines here are that (i) your writeup must make clear how you are adding at-least-equivalent value to what the above rubric prescribes, and (ii) you must integrate your experience with your group in a way that allows the others to "put their oars in," even as critics—or as potential later teammates.