# IDS 2935: AI, Philosophy, and Society
# Quest 1

## I. General Information

### Meeting Times and Location

- Monday/Wednesday 4:05-4:55 PM
- Computer Science and Engineering Building E222
- This course has a required discussion section (meeting times vary).

### Instructor

- David Michael Grant
- Griffin-Floyd Hall 330
- Office hours: Wednesdays 11:30-12:30 PM and by appointment
- david.grant@ufl.edu

### Teaching Assistant(s)

- John ("Jack") Madock
- Griffin-Floyd Hall 316
- Office hours: Wednesdays 11:30-2:30 PM
- johnmadock@ufl.edu

### Course Description

In the past few years, the capabilities of AI-based systems have grown explosively due to the development of a new technology, *large language models*. These systems, known informally as "chatbots," are trained on a significant portion of the text and images that humanity has collectively produced over centuries. As a result, they have developed the ability to perform tasks that we normally associate with human-level intelligence, such as writing essays, writing computer programs, and passing graduate-level exams. In this course, we will explore the philosophical and social implications of this powerful new technology. Are chatbots intelligent in the same sense that we are intelligent? Will they take jobs previously held by highly skilled human workers, such as lawyers, doctors, and software engineers? Could we fall in love with chatbots, and could they fall in love with us? Why are so many of the experts developing chatbots concerned that they might destroy humanity? What can we do to *stop* them from destroying humanity? Will we one day be able to "upload" our minds to computer servers, in effect becoming chatbots ourselves?

As we explore these questions, we will engage with research from several academic disciplines, including computer science, psychology, philosophy, and economics. Assignments will focus on original research into existing AI-based technologies as well as critical reflection on how we want AI to shape society going forward.

This course provides both Quest 1 and General Education Humanities credit. As a Quest 1 course, it focuses on multidisciplinary exploration of an essential question about the human condition (how AI is changing society, and how should we, as citizens, respond?) and emphasizes both experiential learning and self-reflection. As a General Education Humanities course, it familiarizes students with analytical tools from several humanities disciplines in order to equip them with the ability to approach this question in a rigorous way and from multiple theoretical perspectives.

## Quest and General Education Credit

- Quest 1
- Humanities

*This course accomplishes the [Quest](Quest) and [General Education](General Education) objectives of the subject areas listed above. A minimum grade of C is required for Quest and General Education credit. Courses intended to satisfy Quest and General Education requirements cannot be taken S-U.*

## Required Readings and Works

No required materials or fees. All readings and other course materials will be posted to Canvas.

# II. Graded Work

## Description of Graded Work

|  | Description | Points |
|---|---|---|
| **Perusall** | Annotated weekly readings in Perusall | 100 |
| **Group exercises** | Five group activities focused on experiential learning and reflection | 200 |
| **Short essays** | Two short analytical essays | 300 |
| **Capstone project presentation** | Group presentation exploring a philosophical question raised by the course | 100 |
| **Capstone project essay** | Short analytical essay exploring the philosophical question selected by the student's group | 200 |
| **Participation** | Participation in lecture and section | 100 |

### Perusall (100 points)

Collaborative annotation of weekly readings on the Perusall online collaborative annotation platform. You will be asked to annotate approximately ten readings in total. Your grade will be determined by the quantity and quality of your participation in these collaborative discussions over the course of the semester.

Note that while many of the readings for this course are short, they are also very dense—you will need to read them *carefully* and *multiple times* in order to fully absorb the material. I will provide instructions about how to get the most out of the readings in lecture.

Perusall annotations are required and will not be accepted late.

- You can skip one Perusall annotation assignment without penalty.
- Each subsequent Perusall annotation assignment that you fail to submit will result in a 20% deduction from your Perusall grade.

You are responsible for doing all assigned readings, regardless of whether there is an associated Perusall assignment.

## Group exercise reports (5 x 40 points = 200 points)

Five small group exercises focused on experimenting with contemporary AI systems or reflecting on the role AI plays (or should play) in our society. The exercises will be completed in section and graded on the basis of a brief written report. Each member of the group will submit their own version; the group will then select which version to submit for grading. Group members will normally receive the same grade, based on the overall quality of the report submitted by the group (see the rubric for written assignments below). However, students that fail to contribute adequately (based on peer evaluations) may be required to re-do the assignment for an individualized grade. Students that do not submit an individual version of the report will receive a zero for the assignment.

## Short papers (1 x 100 + 1 x 200 = 300 points)

Two thesis-driven analytical essays (1000-1250 words each) on assigned topics relating to the material read and discussed in class. Each essay will ask you to develop and defend your own position on a question posed by the readings and/or lectures for the course.

## Capstone project presentation (100 points)

Students will form small groups and select a philosophical question raised by the course to explore in greater detail. Groups will work together to formulate an answer to their chosen philosophical question, and develop a short presentation articulating and defending that answer. Groups will then deliver their presentations in lecture during the last two weeks of class. Groups will normally receive a single grade for the presentation, except in cases where one or more group members fails to adequately contribute.

## Capstone project essay (200 points)

Prior to delivering their group's capstone project presentation, students will submit a 750-word analytical essay articulating and defending an answer to the group's chosen philosophical question. This answer may be, but need not be, the same as the answer defended by the group. Essays should be informed by the group's discussions, but must be written individually rather than in collaboration with other group members. Students will then be asked to revise this paper in the light of the feedback they receive from the class during their group's presentation. The deadline for submitting a revised essay will be the Monday following the presentation. Only the final version of the essay will be graded. However, students must submit a complete draft to receive credit for the assignment.

## Participation (100 points)

10% of your final grade will be determined by the quantity and quality of your in-class participation in the course. You will receive a letter grade for participation representing the overall quality of your participation over the course of the semester. Below is a summary of course standards for different letter grades (borderline cases will receive "+" or "−" grades as appropriate):

- A (100%): Outstanding participation. Highly engaged during lecture and section, with active participation in small group discussions and frequent constructive contributions to full-class

discussions. In-class contributions reflect excellent understanding of and critical engagement with the content of lectures and assigned readings.

- B (86%): Good participation. Moderately engaged during lecture and section, with active participation in small group discussions and occasional constructive contributions to full-class discussions. In-class contributions reflect significant understanding of and critical engagement with the content of lectures and assigned readings.
- C (76%): Minimally adequate participation. Minimally engaged during lecture and section. In-class contributions are infrequent and reflect limited understanding of and critical engagement with the content of lectures and assigned readings.
- D (66%): Poor participation. Attentive during lecture and section, but participating only very infrequently in small-group and full-class discussions. In-class contributions do not reflect understanding of or critical engagement with the content of lectures and assigned readings.
- F (0%): Inadequate participation. Not attentive during lecture and section and/or not participating in either small-group or full-class discussions.

## Grading Scale

For information on how UF assigns grade points, visit: https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/

| A  | 94 – 100% |  | C  | 74 – 76% |
|----|-----------|--|----|----------|
| A- | 90 – 93%  |  | C- | 70 – 73% |
| B+ | 87 – 89%  |  | D+ | 67 – 69% |
| B  | 84 – 86%  |  | D  | 64 – 66% |
| B- | 80 – 83%  |  | D- | 60 – 63% |
| C+ | 77 – 79%  |  | E  | <60      |

## Grading Rubrics

### Writing Rubric

(Applies to all written work)

|  | **Excellent** | **Good** | **Fair** | **Poor** |
|---|---|---|---|---|
| **Content** | Fully addresses all aspects of the prompt. Demonstrates excellent understanding of and critical engagement with course material. Uses appropriate details and citations from the text(s).<br><br>(5-6 points) | Fully addresses all aspects of the prompt. Demonstrates good understanding; minor points may be misstated or omitted. Demonstrates good critical engagement with course material. Uses appropriate details and citations from the text(s).<br><br>(3-4 points) | Significant misunderstandings of course material. Key ideas are explained incorrectly or omitted. Limited critical engagement with course material. Little textual evidence to back up interpretation of the readings.<br><br>(1-2 points) | Very substantial misunderstandings. Key ideas are explained incorrectly or omitted. Very limited evidence of critical engagement with course material. Textual evidence is missing or not relevant.<br><br>(0 points) |
| **Writing** | Written in simple, non-technical language. Develops a small number of points thoroughly. Well-organized and easy to follow. Few, if any, spelling or grammatical errors.<br><br>(4 points) | Mostly clear, focused, and organized, but includes some confusing passages, digressions/repetitive passages, or lapses in organization. Some spelling or grammatical errors.<br><br>(3 points) | Significant lapses in clarity, focus, or organization. Significant spelling or grammatical errors that interfere with understanding.<br><br>(1-2 points) | Very difficult to understand; many spelling or grammatical errors.<br><br>(0 points) |

| | |
|---|---|
| 10 points | A+ (100%) |
| 9 points | A (96%) |
| 8 points | A- (93%) |
| 7 points | B+ (89%) |
| 6 points | B (86%) |
| 5 points | B- (83%) |
| 4 points | C+ (79%) |
| 3 points | C (76%) |
| 2 points | C- (73%) |
| 1 points | D (66%) |
| 0 points | F (60%) |

# Project Presentation Rubric

(Applies to the in-person presentation component of the capstone project)

|  | Excellent | Good | Fair | Poor |
|---|---|---|---|---|
| **Content** | Fully addresses all aspects of the prompt. Demonstrates excellent understanding of and critical engagement with course material.

(5-6 points) | Fully addresses all aspects of the prompt. Demonstrates good understanding of course material with minor errors or omissions. Demonstrates substantial critical engagement with course material.

(3-4 points) | Significant misunderstandings of course material. Key ideas are explained incorrectly or omitted. Limited critical engagement with course material.

(1-2 points) | Very substantial misunderstandings. Key ideas are explained incorrectly or omitted. Very limited evidence of critical engagement with course material.

(0 points) |
| **Presentation** | Ideas are presented in simple, non-technical language. Develops a small number of points thoroughly. Well-organized and easy to follow. Polished and engaging. Presentational materials (e.g., slides) are polished and effective.

(4 points) | Mostly clear, focused, organized, and polished. Brief segments of the presentation are confusing, repetitive, or off-topic. Uses effective presentational materials.

(3 points) | Significant lapses in clarity, focus, or organization. Presentational materials are ineffective or poorly constructed.

(1-2 points) | Very difficult to understand. Presentational materials are absent or incomplete.

(0 points) |

10 points     A+ (100%)
9 points      A (96%)
8 points      A- (93%)
7 points      B+ (89%)
6 points      B (86%)
5 points      B- (83%)
4 points      C+ (79%)
3 points      C (76%)
2 points      C- (73%)
1 points      D (66%)
0 points      F (60%)

# III. Tentative schedule and readings

Please note that the following is tentative and subject to change.

Please see Canvas for an up-to-date schedule and list of readings.

## Unit 1: Intelligent machines

Wed 08/23 · L1 · What are the machines up to?

- No readings

Mon 08/28 · L2 · How smart are machines getting?

- The Ezra Klein Show, "A.I. could solve some of humanity's hardest problems. It already has" (podcast)

Wed 08/30 · L3 · How do we tell if a machine is smart?

- Turing, "Computing machinery and intelligence"

Mon 09/04 · Holiday (no class)

Wed 09/06 · L4 · Are machines already smart?

- Bubeck et al., "Sparks of Artificial General Intelligence: Early experiments with GPT-4"

Mon 09/11 · L5 · Maybe machines aren't so smart?

- Marcus and Davis, "GPT-3, Bloviator: OpenAI's Language Generator Has No Idea What It's Talking About"
- Hofstadter, "The Shallowness of Google Translate"

## Unit 2: Social machines

Wed 09/13 · L6 · Will we lose our jobs to machines?

- DePillis and Lohr, "Tinkering with ChatGPT, workers wonder: Will this take my job?"
- Gheaus and Herzog, "The goods of work (other than money!)"

Mon 09/18 · L7 · If machines do it better, why bother?

- Karlan, "Human achievement and artificial intelligence"

Wed 09/20 · L8 · Can we fall in love with machines?

- HiPhi Nation, "Love in the time of Replika" (podcast)

Mon 09/25 · L9 · Can machines fall in love with us?

- Frank and Nyholm, "From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?" (pages 219–238)

Wed 09/27 · L10 · Should machines kill?

- Satariano et al., "Killer Robots Aren't Science Fiction. A Push to Ban Them Is Growing."

Mon 10/02 · L11 · When a machines kills, who should we blame?

- Sparrow, "Killler robots"

Wed 10/04 · L12 · Will machines destroy us?

- The Future of Humanity Institute, "Pause Giant AI Experiments: An Open Letter"
- Bostrom, "Existential Risk Prevention as Global Priority"
- Srinivasan, "Stop the Robot Apocalypse"

## Unit 3: Moral machines

Mon 10/09 · L13 · Why don't machines do what we want?

- Christian, The Alignment Problem, chapter 1

Wed 10/11 · L14 · Why is to hard to get machines to do what we want?

- Amodei et al, "Concrete Problems in AI Safety"

Mon 10/16 · L15 · What should we want machines to do?

- Karen Hao, "Should a Self-Driving Car Kill the Baby or the Grandma? Depends on Where You're From"
- Abby Jaques, "Why the Moral Machine is a Monster" (pages 1–9)

Wed 10/18 · L16 · How can we make machines do what we want?

- Asimov, "The three laws"
- Anderson and Anderson, "Creating an Ethical Intelligent Agent"

Mon 10/23 · L17 · Why not just ask machines to do what we want?

- Hard Fork, "Dario Amodei, C.E.O. of Anthropic, on the Parodoxes of A.I. Safety" (podcast) (Amodei interview only)

Wed 10/25 · L18 · Would a moral machine act like a moral human?

- Grant, "Ethics for Artificial Agents"

## Unit 4: Superintelligent machines

Mon 10/30 · L19 · Will machines get smarter than us?

- Chalmers, "The singularity: a philosophical analysis" (pp. 1-22)

Wed 11/01 · L20 · If machines get smarter than us, can we control them?

- Ex Machina (film)
- Chalmers, "The singularity: a philosophical analysis" (pp. 22-34)

Mon 11/06 · L21 · Can we become machines?

- Black Mirror, "Be right back" (episode)

Wed 11/08 · L22 · How could we become machines?

- Chalmers, "The singularity: a philosophical analysis" (34-54)

## Unit 5: Final project

Mon 11/13 · L23 · Review and final project orientation

- No new readings

Wed 11/15 · L24 · Work on final project

- No new readings

Mon 11/20 · L25 · Work on final project

- No new readings

Wed 11/22 · Holiday (no class)

Mon 11/27 · L26 · Project presentations

- No new readings

Wed 11/29 · L27 · Project presentations

- No new readings

Mon 12/04 · L28 · Project presentations

- No new readings

Wed 12/06 · L29 · Project presentations

- No new readings

# IV. Student Learning Outcomes (SLOs)

At the end of this course, students will be expected to have achieved the Quest and General Education learning outcomes as follows:

**Content***: Students demonstrate competence in the terminology, concepts, theories and methodologies used within the discipline(s).*
- Identify, describe, and explain how the resources available in the humanities and sciences can help us understand artificial intelligence's present and future effects on society (Content SLOs for Gen Ed Humanities and Q1). Assessed by Group Exercises, Short Papers, Capstone Project, Participation, and Perusall.
- Identify, describe, and explain how the resources available in the humanities and sciences can help us reflect on how we *should* want artificial intelligence to affect society (Content SLOs for Gen Ed Humanities and Q1). Assessed by Group Exercises, Short Papers, Capstone Project, Participation, and Perusall.

**Critical Thinking**: *Students carefully and logically analyze information from multiple perspectives and develop reasoned solutions to problems within the discipline(s).*
- Identify, analyze, and evaluate themes in public and scholarly discussions of artificial intelligence and its effects on society (Critical Thinking SLOs for Gen Ed Humanities and Q1). Assessed by Group Exercises, Short Papers, Capstone Project, Participation, and Perusall.
- Analyze and evaluate the particular questions about artificial intelligence and its effects on society that we discuss in the course, such as questions about how artificial intelligence will affect our work and romantic lives (Critical Thinking SLOs for Gen Ed Humanities and Q1). Assessed by Group Exercises, Short Papers, Capstone Project, Participation, and Perusall.

**Communication**: *Students communicate knowledge, ideas and reasoning clearly and effectively in written and oral forms appropriate to the discipline(s).*
- Develop and present clear and effective responses to essential questions about how artificial intelligence is, will, and should affect society in oral and written forms appropriate to the relevant humanities disciplines incorporated into the course (Communication SLOs for Gen Ed Humanities and Q1). Assessed by Group Exercises, Short Papers, Capstone Project, Participation, and Perusall.

**Connection**: *Students connect course content with meaningful critical reflection on their intellectual, personal, and professional development at UF and beyond.*
- Reflect on how artificial intelligence is affecting us and our communities (Connection SLO for Q1). Assessed by Group Exercises 2, 4, and 5, Capstone Project, Participation, and Perusall.
- Reflect on what we value, and how those values should shape the way we interact with AI-based technologies and the way we respond to AI's social effects (Connection SLO for Q1). Assessed by Group Exercises 2, 4, and 5, Capstone Project, Participation, and Perusall.

# V. Quest Learning Experiences

### 1. Details of Experiential Learning Component

Experiential learning is an important part of this course and will be incorporated into several assignments, including the group exercises, short analytical essays, and capstone project. These assignments will involve experimenting with large language models and conducting original philosophical research.

### 2. Details of Self-Reflection Component

Throughout the course, you will have the opportunity to reflect on how AI-based technologies have affected your own life and the lives of those around you, whether you believe those effects are good or bad (and why), and how we should address the social changes that AI is already bringing about and may bring about in the future. Time for self-reflection will be integrated into nearly every lecture, and will also be incorporated into the group exercises.

# VI. Required Policies

### Attendance Policy

Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at:
https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx

### Students Requiring Accommodation

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

### UF Evaluations Process

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

## University Honesty Policy

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/ ) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

## Counseling and Wellness Center

Contact information for the Counseling and Wellness Center: http://www.counseling.ufl.edu/ , 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

## The Writing Studio

The writing studio is committed to helping University of Florida students meet their academic and professional goals by becoming better writers. Visit the writing studio online at http://writing.ufl.edu/writing-studio/ or in 2215 Turlington Hall for one-on-one consultations and workshops.

## In-Class Recordings

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the

publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.