

Philosophy of Artificial Intelligence (PHI 4930)

COURSE DESCRIPTION

In this course, we will explore fundamental questions about the nature, possibility, and consequences of artificial intelligence. What is intelligence, and could a computer be intelligent in the same way you and I are intelligent? What about consciousness—could something made out of computer chips know what it is like to see the color red, or hear beautiful music? Do we have moral obligations to machines? What do computer scientists mean when they describe a machine learning system as a “black box”? (Are humans “black boxes” in the same sense?) Why do many people believe that it is morally wrong to deploy lethal autonomous weapons in war, and are they right? As we discuss these questions, we will engage with both classic and contemporary readings in the philosophy of artificial intelligence.

MEETING TIMES AND LOCATION

Tuesdays 11:45 AM–1:40 PM and Thursdays 12:50 PM–1:40 PM in Matherly 51

INSTRUCTOR

David Gray Grant, Ph.D.
Assistant Professor of Philosophy
University of Florida
david.grant@ufl.edu
Office hours: TBD
Office location: 330 Griffin-Floyd Hall

REQUIRED MATERIALS

All materials will be posted on Canvas.

COURSE OBJECTIVES

1. Explain and evaluate arguments about whether artificial general intelligence is possible.
2. Explain and evaluate arguments about ethical challenges raised by artificial intelligence.
3. Discern the structure of arguments, represent them fairly and clearly, and evaluate them for cogency.
4. Formulate original arguments, anticipate objections, and respond in a conscientious fashion.
5. Read and discuss complex philosophical texts.
6. Speak and write clearly and persuasively about abstract and conceptually elusive matters.

COURSE WEBSITE

This course is supplemented by online content in the Canvas e-Learning environment. PDF readings, an electronic copy of the syllabus, and assignment submission portals can be found on the course website.

- To login to the e-Learning site for this course, go to <https://lss.at.ufl.edu/>, click the e-Learning in Canvas button, and on the next page enter your Gatorlink username and password. You can then access the course e-Learning environment by selecting PHI 3681 from the Courses pull-down menu at the top of the page.
- If you encounter any difficulties logging in or accessing any of the course content, contact the UF Computing Help Desk at (352) 392-4537 or <http://helpdesk.ufl.edu>.
- Please do not contact the course instructor regarding computer issues (I am unlikely to be able to help you!).

COMMUNICATION POLICY

Announcements

Course announcements will be posted on Canvas. You are responsible for checking Canvas at least once a week to make sure that you do not miss important announcements.

Contacting Dr. Grant

Please feel to reach out to me directly by email (david.grant@ufl.edu) if you have any questions (or would just like to chat about the course).

- I make an effort to respond to email from students within two (2) business days. Note that emails do sometimes get lost – due to spam filtering, for instance. Please send me another email or come up to me after class if you do not hear back within two business days (I promise I am not ignoring you intentionally!).

ASSESSMENT

Two papers (70%)

You will write two papers in response to prompts that will be provided on Canvas. The first paper will be 1500–2000 words in length, and will be worth 30% of your final grade. The second will be 2200–3000 words in length, and will be worth 40% of your final grade.

Participation (30%)

30% of your final grade will be determined by the quantity and quality of your participation in the course. Activities that will contribute to your participation grade include:

- submitting weekly reading responses;
- regularly attending lecture;
- participating in small-group and full-class discussions during lecture; and

- submitting occasional small ad hoc assignments (announced in class or on Canvas).

Participation grade

You will receive a letter grade for participation representing the overall quality of your participation in the above activities 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: Outstanding participation in all categories. Highly engaged during lecture, with active participation in small group discussions and frequent constructive contributions to full-class discussions. Reading responses, ad hoc assignments, and in-class contributions reflect excellent understanding of and critical engagement with the content of lectures and assigned readings.
- B: Good participation in all categories. Moderately engaged during lecture, with active participation in small group discussions and occasional constructive contributions to full-class discussions. Reading responses, ad hoc assignments, and in-class contributions reflect significant understanding of and critical engagement with the content of lectures and assigned readings.
- C: Minimally adequate participation in all categories. Minimally engaged during lecture. Reading responses, ad hoc assignments, and in-class contributions reflect limited understanding of and critical engagement with the content of lectures and assigned readings.
- D or F: Inadequate participation in one or more categories.

Attendance requirement

Lecture attendance is required. Repeat unexcused absences will be subject to a penalty.

- You are allowed three unexcused absences without penalty.
- Each unexcused absence beyond the first three will result in a 10% deduction from your participation grade (and therefore 2% of your final grade in the course). For example, if your participation grade based on the above standards was an 86 (‘B’) and you had four unexcused absences, your final participation grade in the course would be a 76 (‘C’).
- If you need to miss lecture for any reason, please contact me as soon as possible so that we can discuss whether the absence will be excused and what sort of documentation will be necessary (if applicable). If you do not contact me about an absence within one week following the absence, it will not be excused except under extraordinary circumstances.
- Poor conduct in class will also result in a 10% deduction. Examples of poor conduct include arriving late, disruptive behavior, falling asleep, failing to participate in small group activities, sending text messages, checking email, etc.

Note that the “Roll Call Attendance” grade that Canvas automatically generates is not directly related to your Participation grade in the class.

Reading responses

Each Monday (by 11:59 PM), you will need to turn in a short, one-page (about 500 words) response to one of the readings for the week (your choice). The reading response should try to clearly summarize one of the

main arguments or claims of the reading you have chosen and engage with it critically in some way. Further instructions will be posted on Canvas.

- You can skip two reading responses total without penalty.
- Each subsequent reading response that you fail to submit will result in a 10% deduction from your participation grade.

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

To receive a given grade, you must meet or exceed the lower threshold specified. See <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/> for more information about UF grading policies.

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.**

#	Date	Topic	Readings and assignments
1	1/9	Introduction	No readings
2	1/11	Intelligent agents	Russell and Norvig, <i>Artificial Intelligence: A Modern Approach</i> , Chapter 2 (“Intelligent Agents”)
3	1/16	The Turing test	Turing, “Computing Machinery and Intelligence” Kissinger, Schmidt and Huttenlocher, “The Challenge of Being Human in the Age of AI”
4	1/18	The Turing test	No new readings
5	1/23	The Turing test	Block, “Psychologism and Behaviorism,” pp. 5-22

6	1/25	The frame problem	Dennett, "Cognitive wheels: The frame problem in AI," pp. 181-192
7	1/30	Has the frame problem been solved?	Marcus and Davis, "GPT-3, bloviator: OpenAI's language generator has no idea what it's talking about" Chalmers, "GPT-3 and General Intelligence" Askill, "GPT-3: Towards Renaissance Models" Heaven, "ChatGPT is OpenAI's latest fix for GPT-3. It's slick but still spews nonsense"
8	2/1	The Chinese room argument	Searle, "Minds, brains, and programs" <i>Paper 1 draft due 2/6</i>
9	2/6	The nature of mental states	Putnam, "The nature of mental states"
10	2/8	The nature of mental states	No new readings
11	2/13	Consciousness / writing workshop	Chalmers, "The hard problem of consciousness"
12	2/15	Consciousness/ Are LLMs conscious?	Chalmers, "Are large language models sentient?" <i>Paper 1 due 2/16</i>
13	2/20	Are LLMs conscious?	No new readings
14	2/22	The singularity	Chalmers, "The singularity: a philosophical analysis" (pp. 1-34)
15	2/27	The singularity	Thorstad, "Against the singularity hypothesis"
16	2/29	Machines as moral patients	Basl, "Machines as moral patients we shouldn't care about (yet)"
17	3/5	Consequentialism crash course	Shafer-Landau, "Consequentialism"
18	3/7	The singularity and personal identity	Chalmers, "The singularity: a philosophical analysis" (34-54)
<i>3/11-3/19 Spring Break</i>			
19	3/19	The black box problem	Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms"
20	3/21	The right to explanation	Vredenburg, "The right to explanation"
21	3/26	The right to explanation	No new readings

22	3/28	Computational transparency	Creel, "Transparency in complex computational systems"
23	4/2	Ethics of LAWS	No new readings
24	4/4	Responsibility gaps	Sparrow, "Killer machines"
25	4/9	Responsibility gaps	Boonin, "Autonomous Weapon Systems and the Problem of Responsibility" (excerpts)
26	4/11	Responsibility gaps	No new readings
27	4/16	LAWs and right reasons	Purves et al., "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons"
28	4/18	Writing workshop	No new readings
29	4/23	LAWs and right reasons	No new readings <i>Final paper due 4/28</i>

ADDITIONAL INFORMATION

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-conducthonor-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.

Academic Dishonesty

All sources and assistance used in preparing your papers and presentations must be precisely and explicitly acknowledged. Please review the guidelines on attribution and plagiarism found here: <https://guides.uflib.ufl.edu/copyright/plagiarism>. The web creates special risks here. Cutting and pasting even a few words from a web page or paraphrasing material without a reference constitutes plagiarism. If you are not sure how to refer to something you find on the internet, you can always give the URL.

Students are not permitted to use large language models (such as ChatGPT) or other forms of generative AI to complete written assignments for this course. Submitting work produced using one of these tools without prior written authorization constitutes academic dishonesty.

You should expect the minimum penalty for academic dishonesty to be a grade of E for the class (not just the assignment). All incidents of academic dishonesty will be reported to Student Judicial Affairs. Repeat offenders may be penalized by suspension or expulsion from the university.

Students with Disabilities

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center. [Click here to get started with the Disability Resource Center](#). 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.

Classroom Conduct

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student’s legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records.

Laptop and Cellphone Policy

Laptop and cellphone use is prohibited in this class. Please turn off your phone as soon as you enter the classroom. First-time violations will receive a verbal warning. Subsequent violations will yield a loss of participation points.

Attendance, Illness, Religious Holidays, and Twelve Day Rule

Requirements for class attendance, religious holidays, 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>.

Excerpt from the statement on absences:

“Absences count from the first class meeting. In general, acceptable reasons for absence from or failure to participate in class include illness, serious family emergencies, special curricular requirements (e.g., judging trips, field trips, professional conferences), military obligation, severe weather conditions, religious holidays, and participation in official university activities such as music performances, athletic competition or debate. Absences from class for court-imposed legal obligations (e.g., jury duty or subpoena) must be excused. Other reasons also may be approved.”

Course evaluation 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/>.”

Campus Resources

Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit U Matter, We Care website to refer or report a concern and a team member will reach out to the student in distress.

Counseling and Wellness Center: Visit the Counseling and Wellness Center website or call 352-392-1575 for information on crisis services as well as non-crisis services.

Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need, or visit the Student Health Care Center website.

University Police Department: Visit UF Police Department website or call 352-392-1111 (or 9-1-1 for emergencies).

UF Health Shands Emergency Room / Trauma Center: For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; Visit the UF Health Emergency Room and Trauma Center website.

GatorWell Health Promotion Services: For prevention services focused on optimal wellbeing, including Wellness Coaching for Academic Success, visit the GatorWell website or call 352-273-4450.

Academic Resources

E-learning technical support: Contact the UF Computing Help Desk at 352-392-4357 or via e-mail at helpdesk@ufl.edu.

Career Connections Center: Reitz Union Suite 1300, 352-392-1601. Career assistance and counseling services.

Library Support: Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center: Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420. General study skills and tutoring.

Writing Studio: 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers.

Student Complaints On-Campus: Visit the Student Honor Code and Student Conduct Code webpage for more information.

On-Line Students Complaints: View the Distance Learning Student Complaint Process.