

Philosophy 3930: Philosophy of Artificial Intelligence Syllabus

PHI 3930; Section 12E0/12E1

Spring 2021

Tuesdays 11:45am - 1:40pm, Matherly Hall 113

Thursdays 12:50pm - 1:40pm, Matherly Hall 114

Instructor Information

Chris Dorst

cdorst@ufl.edu

Office: 312 Griffin-Floyd Hall

Office Hours (Zoom): Wednesdays 12:00pm - 3:00pm, or by appointment

Phone (office): (352) 273-1813

Course Description

This class will explore some fundamental questions about the nature, possibility, and potential consequences of artificial intelligence. We will begin by considering some questions about intelligence in general: What is its nature? How is it measured? Is it an objective characteristic? We will then turn to the issue of whether it is possible to create genuine *artificial* intelligence, and if so, how it might be done: Could a computer truly be said to *think* and *understand*? And what methods might we use to develop an intelligent computer? Finally, we will consider the potential societal consequences of the development of a genuine artificial intelligence: Could we reasonably hope to control it? What are the potential benefits and detriments? What is the likelihood of an “intelligence explosion,” and what might be the effects if one occurs? In this last unit, we will work through Nick Bostrom’s recent book, *Superintelligence*, and critically evaluate it on the basis of our earlier topics.

Learning Objectives

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

- Explain what we presently understand about the property of general intelligence
- Explain how general intelligence is measured, and how intelligence tests are designed
 - Evaluate the philosophical implications of these points
- Develop an informed philosophical opinion about whether genuine artificial intelligence is possible
- Explain the ways in which we might either recreate or mimic genuine intelligence in a computer
- Explain and evaluate arguments about the potential ramifications of the development of genuine artificial intelligence
- Write persuasively in support of their positions

Academic Honesty

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 (<http://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.

Plagiarism on any assignment will automatically result in a grade of "E" for the course. Plagiarism is defined in the University of Florida's Student Honor Code as follows: "A student shall not represent as the student's own work all or any portion of the work of another. Plagiarism includes (but is not limited to): a. Quoting oral or written materials, whether published or unpublished, without proper attribution. b. Submitting a document or assignment which in whole or in part is identical or substantially identical to a document or assignment not authored by the student." Students found guilty of academic misconduct will be prosecuted in accordance with the procedures specified in the UF honesty policy.

Attendance and Classroom Policies

This is a hybrid course, meaning that some students are enrolled for face-to-face instruction, and others are enrolled for online Zoom meetings.

- If you are enrolled in the online section, you must attend virtually instead of in-person. The Zoom meeting room URL will be posted on Canvas the week before classes start.
- If you are enrolled in the in-person section, you are permitted to attend the class in-person. That said, we will need to be flexible and ready to quickly transition to a fully-online course depending on the state of the pandemic. In-person meetings will occur in Matherly Hall 113 (on Tuesdays) and Matherly Hall 114 (on Thursdays).
 - When you attend in-person, please make sure that you are wearing a mask, and that you bring a laptop or tablet so that you may participate in any of the Zoom meeting discussions. (See the “COVID-19-Related Practices...” section below.)
- For the in-person section, 51% of class meetings must occur in person. My proposal, which I will discuss with the in-person students, is that our default meetings will be online on Tuesdays and in-person on Thursdays. To hit the minimum of 51% in-person, we will need to schedule a few Tuesday meetings in-person later in the semester, which will be done at the appropriate time. I intend to remain flexible about this arrangement in accord with the needs and wants of students enrolled in the in-person section, while adhering to CLAS requirements and guidelines.

Respectful participation is expected of everyone. Each class has assigned readings, which will require critical engagement and reflection. You should come to class prepared to discuss these readings and demonstrate that you have thought critically about them beforehand.

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>.

COVID-19-Related Practices (for In-Person Students)

In response to COVID-19, the following policies and requirements are in place to maintain your learning environment and to enhance the safety of our in-classroom interactions.

- You are required to wear approved face coverings at all times during class and within buildings. Following and enforcing these policies and requirements are all of our responsibility. Failure to do so will lead to a report to the Office of Student Conduct and Conflict Resolution.
- This course has been assigned a physical classroom with enough capacity to maintain physical distancing (6 feet between individuals) requirements. Please utilize designated seats and maintain appropriate spacing between students. Please do not move desks or stations.
- Sanitizing supplies are available in the classroom if you wish to wipe down your desks prior to sitting down and at the end of the class.
- Follow instructor guidance on how to enter and exit the classroom. Practice physical distancing to the extent possible when entering and exiting the classroom.
- If you are experiencing COVID-19 symptoms ([Click here for guidance from the CDC on symptoms of coronavirus](#)), please use the UF Health screening system and follow the instructions on whether you are able to attend class. [Click here for UF Health guidance on what to do if you have been exposed to or are experiencing Covid-19 symptoms](#).
 - Course materials will be provided to you with an excused absence, and you will be given a reasonable amount of time to make up work. [Find more information in the university attendance policies](#).

Canvas e-Learning Environment

This course is supplemented by online content in the e-Learning environment known as "Canvas." 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 2010 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. Please do not contact the course instructor regarding computer issues.

Online Course Evaluation

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/>.

Accommodation for 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.

Health and Wellness Resources

- *U 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](#).

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 Student Complaints*: [View the Distance Learning Student Complaint Process](#).

Course Text (Required)

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Grading Breakdown

Grades are assigned consistent with [current UF grading policies](#). The breakdown is as follows:

- Quizzes, etc.: 20%
- Paper 1: 20%
- Paper 2: 20%
- Midterm Exam: 20%

- Final Exam: 20%

The following is a brief description of each of these assignment categories.

- **Quizzes, etc.:** To prepare for our discussions, we need to study the readings carefully. Try to outline what the author's main claim is, what their principal reasons for that claim are, what alternative position(s) they consider, and the reason(s) why they do not adopt those alternatives. To reward you for doing this work, there will be a number of quizzes during the course of the semester. Each quiz will contain several true-false or multiple choice questions about the readings. Typically they will be assigned at the beginning of a week, and be due by the end of the week. Quizzes will be available on Canvas.

The "etc." portion is meant to cover ad hoc assignments that may prove useful as the semester progresses. For example, we may occasionally divide into small groups to discuss a given question or topic, and I will sometimes ask you to come up with a group response. These responses will count toward the "Quizzes, Etc." component of your course grade.

- **Papers:** You will write two papers for this course. For each paper, you will be given the choice of several prompts, of which you must choose one and write a 4-5 page paper in response. The prompts will concern topics we have read about and discussed in class. The papers give you the opportunity to illustrate your understanding of the material, as well as to develop your own views. Late papers will be deducted 1/3 of a letter grade (i.e. a +/- increment) for each day past the due date.
- **Midterm Exam:** The midterm exam is a take-home exam due on Friday, March 19 at 11:59pm (to be submitted on Canvas). The exam prompts will be provided one week before the due date. You will have to choose a subset of them to answer with a short essay. The questions will concern the readings and topics we have discussed over the first half of the semester. The combined length of your answers should be at least 1000 words. While you may discuss the question prompts with other students, *you may not work with anyone else on the preparation of your answers.*
- **Final Exam:** The Final Exam is a take-home exam, due on Tuesday, April 27 at 11:59pm (to be submitted on Canvas). The exam will be similar in format to the midterm: it will consist of around five questions, and you will have to choose a subset of them to answer with a short essay. The questions will concern the readings and topics we have discussed over the course of the entire semester. The combined length of your answers should be at least 1000 words. While you may discuss the question prompts with other students, *you may not work with anyone else on the preparation of your answers.*

Lateness

The policy for late assignments depends on the type of assignment. Late quizzes, papers, and Midterm Exams will be deducted 10% for each day past the deadline. Ad hoc in-class assignments falling under the "Etc." description above will not be accepted late. Likewise, the Final Exam cannot be accepted late. If there is a documented medical or personal reason that you need to submit an assignment late, please let me know.

Course Schedule and Readings

The following is a tentative schedule for the course. Any official changes to the schedule will be announced in class and over email. Note that the schedule is divided into weeks; the readings listed for a given week are to be read before classes that week (aside from Week 1). Readings from our course text are labeled “Bostrom, Chapter XX”. Any readings not found in our course text will be posted on Canvas.

Unit 1: The Concept of Intelligence

Week 1: January 12, 14

- John Wasserman, “A History of Intelligence Assessment”

Week 2: January 19, 21

- John Wasserman, “A History of Intelligence Assessment”
- Ned Block and Gerald Dworkin, “IQ: Heritability and Inequality, Part I”, Sections 1-2

Week 3: January 26, 28

- Ned Block and Gerald Dworkin, “IQ: Heritability and Inequality, Part I”, Sections 3-5, 7-8

Unit 2: The Possibility of Artificial Intelligence

Week 4: February 2, 4

- Alan Turing, “Computing Machinery and Intelligence”

Week 5: February 9, 11

- Daniel Dennett, “Can Machines Think?”
- John Searle, “Minds, Brains, and Programs”

Paper 1 Due by February 12 at 11:59pm (submit on Canvas)

Week 6: February 16, 18

- Daniel Dennett, “The Intentional Strategy and Why it Works”

Week 7: February 23, 25

- Bostrom, Chapter 1: “Past Developments and Present Capabilities”
- Selmer Bringsjord and Naveen Govindarajulu, “Artificial Intelligence”, Section 3

Week 8: March 2, 4

- Bostrom, Chapter 2: “Paths to Superintelligence”
- Bostrom, Chapter 3: “Forms of Superintelligence”

Unit 3: The Consequences of Artificial Intelligence

Week 9: March 9, 11

- Bostrom, Chapter 4: “The Kinetics of an Intelligence Explosion”
- Bostrom, Chapter 5: “Decisive Strategic Advantage”

Midterm Exam due by March 12 at 11:59pm (submit on Canvas)

Week 10: March 16, 18

- Bostrom, Chapter 6: “Cognitive Superpowers”
- Bostrom, Chapter 7: “The Superintelligent Will”

Week 11: March 23, 25

- Bostrom, Chapter 8: “Is the Default Outcome Doom?”
- Bostrom, Chapter 9: “The Control Problem”

Week 12: March 30; April 1

- Bostrom, Chapter 10: “Oracles, Genies, Sovereigns, Tools”

Week 13: April 6, 8

- Bostrom, Chapter 12: “Acquiring Values”
- Bostrom, Chapter 13: “Choosing the Criteria for Choosing”

Paper 2 Due by April 9 at 11:59pm (Submit on Canvas)

Week 14: April 13, 15

- Bostrom, Chapter 14: “The Strategic Picture”
- Bostrom, Chapter 15: “Crunch Time”

Week 15: April 20

- TBD/Makeup Time

Final Exam Due by April 27 at 11:59pm (Submit on Canvas)