COURSE SYLLABUS

****

YEAR COURSE OFFERED: 2025
SEMESTER COURSE OFFERED: Spring
DEPARTMENT: Philosophy
COURSE NUMBER: 3681
NAME OF COURSE: Ethics, Data, and Technology
NAME OF INSTRUCTOR: Cameron Buckner

****

The information contained in this class syllabus is subject to change without notice. Students are expected to be aware of any additional course policies presented by the instructor during the course.

****

Time and location: T/H Period 7 1:55-2:45, TUR L007
Office hours: T/H 12:45-1:45 FLO 330B
Instructor e-mail: cameron.buckner@ufl.edu

Discussion sections:
20535: F Period 5 11:45-12:35 TUR 2349
20536: F Period 7 1:55 – 2:45 MAT 0014
20537: F Period 4 10:40-11:30 ROL 0115
21855: F Period 5 11:45-12:35 Mat 0002
21857: F Period 7 1:55-2:45 MAT 0015

Course Description:

This course will explore philosophical issues surrounding the development and deployment of emerging technologies, focusing especially on technological advances based on "deep learning" techniques in computer science. The primary focus will be on ethical and explanatory questions surrounding the use of these systems, which in just a few years have come to have pervasive effects in our daily lives—despite the fact that our understanding of their philosophical implications remaining rudimentary. Questions we will explore are: in what senses are these systems biased, and when is their bias ethically problematic? Can we explain the workings of these vastly complex systems—containing billions of parameters and trained on Internet-scale datasets—in a way that answers to our existing scientific, legal, and ethical practices? Who is responsible when these systems err? And finally: how can humanity adjust to the radical changes these systems are bringing to our social, political, and economic lives without losing our fundamental humanity, and can these systems be designed in a way to align with human values—as opposed to alien, machine objectives?

Please note that all lecture notes and reading responses for this course are available on the course's Canvas site.

Course objectives:

1. Understand the basic architectures and training methods used in contemporary "deep" machine learning research, the differences between this approach and earlier methods in artificial intelligence, and their current applications to software systems in daily life.

2. Develop a basic vocabulary for understanding and ethically evaluating these systems by drawing upon theories and concepts from philosophy.
3. Recognize arguments for various positions in the ethical evaluation of cutting-edge technologies, represent them fairly and clearly, and evaluate their cogency.
4. Formulate original arguments, anticipate objections, and respond to them in a conscientious manner.
5. Read and discuss complex texts from historical sources and contemporary works.
6. Speak and write persuasively on abstract and conceptually difficult issues at the intersection of philosophy and emerging technologies.

Major Assignments:

Weekly Reading Responses	10%
Mid-Term Exam	20%
Group Case Studies (5)	50%
Final Exam	20%

Weekly Reading Responses

Weekly reading response assignments will be posted on Canvas. These will provide a "guided tour" of the readings, helping you navigate philosophical texts by recognizing key claims and arguments. You should complete the reading responses as you go through the reading, by answering true/false, multiple choice, and fill-in-the-blank style questions.

Group Case Studies

In addition to reviewing the topics covered in the main lecture component, most of the time in discussion sections will be spent working on group case studies on the topics of interest. For each case study, you will work on an activity in groups of 4-5, answering questions provided in a prompt. In most cases you will need to interact with an AI tool to address the prompt; in some cases there will be a debate format. You will present your group work to the rest of the class in discussion sections in a 5-10 minute presentation and submit a shared 2-4 page "issue brief" reviewing the work you did as a group. The issue brief should also contain appendices with full transcripts or screenshots of any interactions you had with AI models. Issue briefs will also include a self-evaluation component, where groups evaluate how well they did on the project and what they might do differently if they were to do it over again. Groups are free to divvy up the work as they see fit—some group members might do more writing, others more oral presentation, for example. Grades will be derived from rubrics based on both the presentation and issue brief. 85% of the grade will be shared amongst the whole group, but 15% of each case study grade will be based on peer reviews from your other group mates—so pull your weight.

Students can also opt out of the group work by doing individual research and writing their own longer issue brief, more like a 3-4 page paper. If choosing the individual option, you must also come explain your paper individually and answer questions about it with the TA in office hours. Where possible, I highly encourage all students to choose the group work option, which is likely to be more efficient for everyone and to produce more interesting research.

COURSE SYLLABUS

Exams

Undergraduates will take two in-class essay exams: one mid-term and one final. These will be completed in Blue Books by hand in class. A list of possible topics will be handed out roughly 1 week before the exam date, from which the essay questions will be chosen.

Grading scale

Grading of each assignment will be based on a 100 point scale. Grades for group projects will be derived from the grading template rubrics, which you can see and consider ahead of time. For final letter grades, I reserve the discretion to round up grades near a cutoff line in final averages for students who have reliably turned in assignments and/or improved their performance towards the end of the course.

Percentage greater than or equal to	Letter
93-100 %	A
90-93 %	A-
87-90 %	B+
83-87 %	B
80-83 %	B-
77-80 %	C+
73-77 %	C
70-73 %	C-
65-70 %	D+
60-65 %	D
50-60 %	D-
< 50 %	F

Required Reading

All readings will be posted on the Canvas site.

List of discussion/lecture topics

Students are expected to have read the chapters or articles indicated and to have completed the weekly reading response on Canvas by the assigned day.

Week	Topic	Readings
Week 1 Jan 14	Basic Machine Learning #1: AI & DCNNs	Turing 1950: Computing Machinery and Intelligence Buckner 2018: Deep Learning: A Philosophical Introduction
Week 2 Jan 21	Generative AI and Philosophy of Mind	Buckner 2018: Empiricism without Magic-Transformational Abstraction in DCNNs

COURSE SYLLABUS

Week 3 Jan 28	Basic Machine Learning #2: Transformers	Buckner & Milliere 2022: A Philosophical Introduction to Language Models Pt. 1 Block 1981: Psychologism and Behaviorism *Group Activity #1 Assigned:* Case Study on the Turing Test
Week 4 Feb 4	The Black Box Problem	Rudin 2019: Stop Explaining Black Box Machine Learning Models… Buckner 2023: Black Boxes or Unflattering Mirrors? Comparative Bias in the Science of Machine Behaviour *Group Activity #1 Due:* Case Study on the Turing Test
Week 5 Feb 11	Interventionist Interpretability Methods	Buckner & Milliere 2023: Interventionist Methods for Interpreting Deep Neural Networks Vredenburgh 2022 - "The Right to Explanation" *Discussion Section Group Activity #2 Assigned:* Case Study on Bias in DNNs
Week 6 Feb 18	Algorithmic Bias 1	Fazelpour and Danks 2021: Algorithmic Bias—Senses, Sources, Solutions Julia Angwin 2016 – "Machine Bias" ProPublica Corbett-Davies et al. 2016: "A computer algorithm used for bail…"
Week 7 Feb 25 Feb 27: *Exam #1*	Algorithmic Bias 2 (Madock Lectures)	Johnson 2020 – Algorithmic Bias-on the implicit biases of social technology Creel and Hellman 2022 – The Algorithmic Leviathan *Discussion Section Group Activity #2 Due:* Case Study on Bias in DNNs
Week 8 Mar 4	Responsibility	Mathias 2004 – The Responsibility Gap Tigard 2021 – There is no techno-responsibility Gap

COURSE SYLLABUS

		Discussion Section Group Activity #3 Assigned*:* Debate on Responsibility for AI Accidents
Mar 11	Trust (Madock Lectures)	Simion & Kelp 2023 – Trustworthy Artificial Intelligence Hevelke & Nida-Rumelin 2015 – Responsibility for Crashes of Autonomous Vehicles *Discussion Section Group Activity #3* Due*:* Debate on Responsibility for AI Accidents
Week 9 Mar 18	Spring Break	No Readings
Week 10 Mar 25	The Alignment Problem	Anthropic Team 2022 – Constitutional AI Tubert & Tiehen 2024 – Existential risk and value misalignment *Discussion Section Group Activity #4 Assigned:* Case study on AI Alignment
Week 11 Apr 8	Algorithms, Echo Chambers, and Mental Health	Nguyen 2020 – "Echo Chambers and Epistemic Bubbles" Munroe 2024 – "Echo Chambers, Polarization, and 'Post-Truth'- In Search of a connection" Wells et al. "Facebook knows Instagram is Toxic for Teen Girls" (WSJ) Lewis, "Our minds can be hijacked" (The Guardian) *Discussion Section Group Activity #4 Due:* Case study on AI Alignment
Week 12 Apr 15	Generative AI, Art, and Intellectual Property	Vlaad 2024 – A Portrait of the Artist as Young Algorithm Kieval 2024 – Artificial Achievement Goetze 2024 – AI art is theft *Discussion Section Group Activity #5 Assigned:* Should Generative AI be restrained or embraced?

COURSE SYLLABUS

Week 13 Apr 22	The Future of Work	Danaher 2017 – Will life be worth living in a world without work? Belic 2024 – Institutions, Automation, and Legitimate Expectations *Discussion Section Group Activity #5 Due:* Should Generative AI be restrained or embraced?
Final Exam		5/02/2025 @ 7:30 AM - 9:30 AM

Attendance:

Requirements for class attendance and other work in this course are consistent with university policies that can be found at: https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/

Accommodations:

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the Disability Resource Center. See the "Get Started With the DRC" webpage on the Disability Resource Center site: https://disability.ufl.edu/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.

Grading policies:

Grading policies are consistent with the UFL guidelines found in the catalog here: https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/ .

Evaluation policies:

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/ .

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 Conduct Code specifies a number of behaviors that are in violation of this code and the possible sanctions. See the UF Conduct Code website for more information. If you have any questions or concerns, please consult with the instructor or TAs in this class.

COURSE SYLLABUS

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 education 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 deliver by an 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 presentation 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 guest 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.

Health and Wellness:

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.

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:

COURSE SYLLABUS

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. Call 866-281-6309 or email ask@ufl.libanswers.com for more information.

Teaching Center: 1317 Turlington Hall, 352-392-2010 or to make an appointment 352- 392-6420. General study skills and tutoring.

Writing Studio: Daytime (9:30am-3:30pm): 2215 Turlington Hall, 352-846-1138 | Evening (5:00pm-7:00pm): 1545 W University Avenue (Library West, Rm. 339). Help brainstorming, formatting, and writing papers.

Academic Complaints: Office of the Ombuds; Visit the Complaint Portal webpage for more information.

Enrollment Management Complaints (Registrar, Financial Aid, Admissions): View the Student Complaint Procedure webpage for more information.

Instructional material note:

Instructional materials for this course consist of only those materials specifically reviewed, selected, and assigned by the instructor(s). The instructor(s) is only responsible for these instructional materials.