PHI 3681: ETHICS, DATA, AND TECHNOLOGY Dr. James Gillespie

1. Basic Information

1.1 Instructor Information

Instructor: Dr. James Gillespie (& Dr. Rusong Huang)

• Email: james.gillespie@ufl.edu (& rusonghuang@ufl.edu)

Nota bene: Dr. Huang will lead your Friday discussion sections.

1.2 Course Times and Locations

• Times: TR, 1:55 PM – 2:45 PM (Period 7)

• Location: FLG 270

• Discussion Sections: F, Periods 4, 5, 7 (Led by Dr. Huang)

1.3 Office Hours

Dr. Gillespie's

• M, 12 – 1:30 PM

• W. 12 – 1:30 PM

Griffin – Floyd Hall, FLO 320

Dr. Huang's

- T. 10:30 AM 11:30 PM
- R, 10:30 AM 12:30 PM
- Griffin Floyd Hall, FLO 320

2. Course Overview

2.1 Course Description

In this course, we will explore questions about how emerging technologies should be designed and regulated. What does it mean to say that an algorithm is "biased" against members of a particular social group? Should we be concerned about the fact that technology companies such as Facebook gather vast amounts of data about our online activities? What does it mean to say that a machine learning algorithm is a "black box," and is there something unfair about using such algorithms to decide how to treat people? As we investigate these and other questions about emerging technologies, we will draw on concepts and readings from a variety of different fields, including philosophy, economics, computer science, data science, and law.

2.2 Course Objectives

- Develop a basic vocabulary for discussing the ethical dimensions of data science and its applications.
- Analyze issues and policies concerning emerging "big data" technologies through the application of ethical concepts.
- Critique public policies, social practices, and social institutions that shape, and are shaped by, scientific discovery and technological design.
- Discern the structure of arguments, represent them fairly and clearly, and evaluate them for cogency.
- Formulate original arguments, anticipate objections, and respond in a conscientious fashion.
- Read and discuss complex philosophical texts from both historical sources and contemporary works.
- Speak and write clearly and persuasively about abstract and conceptually elusive matters.

2.3 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 teaching team regarding computer issues (we are unlikely to be able to help you!).

2.4 Required Materials

No book purchases are required, as all readings are made available as PDF files online through the Canvas system. There is a Canvas assignment for each reading.

2.5 Course Structure

This course is broken up into five units as follows:

- Unit I: The Alignment Problem (4 Sections)
- Unit II: Autonomy (3 Sections)
- Unit III: Privacy (4 Sections)
- Unit IV: Fairness (2 Sections)
- Unit V: Transparency (3 Sections)

3. Grading Information

3.1 Grade Breakdown

Attendance and Participation Grade (200 points)	20%
Emerging Technology Case Studies (400 points)	
Midterm Exam (200 points)	
Final Exam (200 points)	
Total	

3.2 Grading Scale

The following grade scale will be used to assign final letter grades for the course. See UF grading policies for assigning grade points at: https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx.

Points	Grade Scale	Grade Value
934–1000	93.4-100 = A	A=4.0
900–933	90.0-93.3 = A-	A-=3.67
867–899	86.7-89.9 = B+	B+=3.33
834–866	83.4-86.6 = B	B=3.00
800–833	80.0-83.3 = B-	B-=2.67
767–799	76.7-79.9 =C+	C+=2.33
734–766	73.4-76.6 =C	C=2.00
700–733	70.0-73.3 = C-	C-=1.67
667–699	66.7-69.9 = D+	D+=1.33
634–666	63.4-66.6 = D	D=1.00
600–633	60.0-63.3 = D-	D-=0.67
0–599	0-59.9 = E	E=0.00

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.

3.3 Attendance and Participation Grade (200 points)

20% 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:

- annotating the weekly readings in Perusall (assignments will be announced on Canvas):
- regularly attending lectures and your assigned discussion section;
- participating in small-group and full-class discussions during lectures and your assigned discussion section; 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 and discussion section, with active participation in small group discussions and frequent constructive contributions to full-class discussions. Perusall annotations, 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 and discussion section, with active participation in small group discussions and occasional constructive contributions to full-class discussions. Perusall annotations, 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 and discussion section. Perusall annotations, 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.

PERUSALL ANNOTATIONS AND AD HOC ASSIGNMENTS. Perusall annotations and *ad hoc* assignments will not be accepted late.

- You can skip two Perusall annotation/ad hoc assignments in total without penalty.
- Each subsequent Perusall annotation/ad hoc assignment that you fail to submit will result in a 10% deduction from your participation grade.

Note: You will need to access each Perusall reading through the link in the associated assignment on Canvas—rather than by opening the Perusall website directly. Otherwise your score will not sync back to Canvas.

INSTRUCTIONS FOR PERUSALL ASSIGNMENTS. For most weeks, I will add one of the required readings to Perusall for you to read and collaboratively annotate with your peers. You can access these through the "Assignments" tab, or by clicking on the assignment listing in the course calendar under "Syllabus."

A few notes:

 See the syllabus for information about how Perusall assignments will affect your grade in the course.

- Effective annotations deeply engage points in the readings, stimulate discussion, offer informative questions or comments, and help others by addressing their questions/confusions.
- You can start a new annotation thread in a Perusall reading assignment by highlighting text or using the figure highlight tool <u>Links</u>, then asking a question or posting a comment. You can also reply to an existing thread or upvote comments/questions posted by other students.
- Try to spread out your annotations—don't just comment on the first few pages!
- More information on how to use Perusall is available on the Perusall website.
- See below for some suggestions about how to approach annotating the readings.

SUGGESTIONS ABOUT WHAT TO WRITE. Below are some helpful suggestions about how to annotate readings from Danny Weltman, a philosophy professor at Ashoka University: "Hopefully you will not have trouble thinking up things to write in Perusall. If you are reading closely and thinking carefully about the texts, probably you will have lots of questions and comments to add. But, if you are having trouble thinking of things to write, here are some options:

- Find an argument that the author is making. Highlight its premises and label them as premises in the argument. Highlight its conclusions and label them as conclusions in the argument.
 - You can also do other things based around arguments, like summarizing an argument in your own words, asking why a certain premise is necessary, asking whether a conclusion follows from some premises, describing how the argument is similar or different to other things in the text or in the course, and so on.
 - You can do this for part of an argument and suggest that other people fill in the other parts of the argument.
- Highlight a confusing word, sentence, or passage, and paraphrase the highlighted material in your own words to make it less confusing.
- Highlight something you find confusing or interesting and ask a question about it.
 - Answer someone's question or add information which you think either might help answer the question or which is related to the question in some other way
 - Note that you also have the question and give more information about why you have this question
- Highlight something that you think is related to another point in the reading, in another reading, in one of the lectures, in one of the reading quizzes, or anything else, and explain that relationship.
- Highlight something you agree or disagree with and say why you agree or disagree
 - Respond to someone's agreement or disagreement by saying whether you concur with them and why you do or do not concur.
- Highlight an important vocabulary word or phrase which the author is using and talk about what it means, perhaps by drawing on the author's definition from an earlier reading or by giving your own understanding of what you think they mean.

- Highlight a principle the author is discussing and give a concrete example of the principle, either hypothetical or from a real life situation.
- Highlight an example the author is discussing and explain what principle they are illustrating with this example.
- Highlight something from the reading which relates to something in the news, or from your life, or any other relevant thing which can help illuminate the reading.

These are not the only things you can do in Perusall, but these are good suggestions if you find you do not have ideas."

ATTENDANCE REQUIREMENT. Attendance in lecture and discussion section is required. Repeat unexcused absences will be subject to a penalty.

- You are allowed four unexcused absences without penalty.
- Each unexcused absence beyond the first four 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 five unexcused absences, your final participation grade in the course would be a 76 ('C').
- If you need to miss a lecture or discussion section for any reason, please contact
 your instructor 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 your instructor 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.

Proper documentation will be required to excuse absences (e.g., a doctor's note). You must provide this documentation to your instructor within one week of the missed class. Arriving at the discussion section more than ten minutes late without prior instructor approval counts as an absence. If you are absent from class for an excused reason, you will not lose any points, and the absence will not exhaust one of your free absences. In most cases, you need documentation of the excuse (see below). Excused absences are limited to the following cases:

- Health. If your physical or mental health causes you to miss class, a note from your doctor or from the Dean of Students within one week of the absence will count as documentation for this.
- Personal Reasons. The Dean of Students will send me a note of excuse if you
 have personal reasons for needing one, such as a death in the family or an
 ongoing medical issue. I need your note from the DOS within one week of the
 absence. Note: I do not get involved in adjudicating good personal reasons from
 bad ones. I let the DOS handle that.
- Some UF Activities can be excused absences. You must provide a note for activities absences prior to the day of absence. Notes may come from the

- activity supervisor, such as a coach. Excused absences will lead to make-up assignments.
- Religious observances. See your instructor at the beginning of the semester about this.

All other absences are typically unexcused, including some very good reasons to miss a class, like a career or family event. I support your interest in these things and understand why you might prioritize them over class, but that does not make them excused absences. Please note that you are responsible for the information presented in class even if you miss class, whether for excused or unexcused reasons.

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

3.4 Emerging Technology Case Studies (400 points)

We will have four discussions of case studies concerning emerging technologies in class. Students will be assigned to discussion groups. Each group member will provide a written analysis of the case study and will use this analysis to inform small group discussions. Groups will then work together to submit a more polished written analysis, which will inform the discussion of the case study during the lecture or discussion section. Polished group analyses will be graded.

3.5 Midterm Exam (200 points)

The midterm exam will be a take-home short essay exam. The questions will concern the readings and topics we have discussed over the first half of the semester. Collaboration on the midterm exam is not permitted; your answers should represent your own understanding of the course material.

3.6 Final Exam (200 points)

Like the midterm exam, the final exam will be a take-home short essay exam. The questions will concern the readings and topics we have discussed over the second half of the semester. Collaboration on the final exam is not permitted; your answers should represent your own understanding of the course material.

4. Course Policies

4.1 COVID-19 Recommendations

In response to COVID-19, the following recommendations are in place to maintain your learning environment, enhance the safety of our in-classroom interactions, and further the health and safety of ourselves, our neighbors, and our loved ones.

- If you are not vaccinated, get vaccinated. Vaccines are readily available and have been demonstrated to be safe and effective against the COVID-19 virus. Visit one uf for screening testing and vaccination opportunities.
- If you are sick, stay home. Please call your primary care provider if you are ill and need immediate care or the UF Student Health Care Center at 352-392-1161 to be evaluated.
- 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.

4.2 Electronic Devices

If you have electronic devices with you—including phones, computers, tablets, and recording devices—these should be silenced, and you will need to suspend interaction with them during class (except to start or stop recording as appropriate). This means you will need to bring a hard copy of your reading and a way to take notes, as you won't be able to use your devices to facilitate in-class activities.

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.

4.3 Classroom Conduct

Philosophy is sometimes mistakenly understood to be a combative exercise. That style does not impress me at all. I encourage you to direct your first effort toward trying to understand and develop both the contributions of the authors we read and the

contributions of others in the class. Where those are limited, of course, good philosophy also happens when you (respectfully) disagree, note tensions, make distinctions, reshape the question, and so on. Students who repeatedly act aggressively in ways that attempt to incite unnecessary conflict or to dominate the discussion and/or other people will receive a zero for participation.

4.4 Academic Honesty

As stated in the UF Student Honor Code [https://sccr.dso.ufl.edu/policies/student-honor-code-studentconduct-code/], all UF students are bound by the following "Honor Pledge"

• The Honor Pledge: We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity by abiding by the Student 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 specifies a number of activities that constitute academic dishonesty as well as the sanctions (that is, penalties) that may result, including suspension and expulsion. The policy in my classes for academic dishonesty is simple:

ZERO TOLERANCE POLICY FOR ACADEMIC DISHONESTY

All incidents are reported to Student Conduct and Conflict Resolution.

CLEAR EVIDENCE OF ACADEMIC DISHONESTY MEANS AN AUTOMATIC FAILING GRADE FOR THE ENTIRE COURSE.

In submitting a report of academic dishonesty, I recommend *at minimum* a failing grade for the course and cite this warning in the syllabus.

If you are at all unsure whether something might violate the code, it is your responsibility to find out. Please consult the code and/or ask the instructor.

Do not be afraid to ask.

In humanities classes, the most common kind of academic dishonesty is plagiarism. In academic writing (including any writing exercise, essay, or paper you write in this class), if you use an idea you got from someone else - whether it's one of the assigned readings, another student, something you saw online, or the like - you are obliged to inform the reader of the source of that idea. Failure to do so is plagiarism.

If you are ever tempted to cheat, please keep in mind that it would be better to get a failing grade for a particular assignment than to get both a failing grade for the entire course and a record of academic dishonesty on file with the UF.

4.5 Accommodations

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the Disability Resource Center [https://disability.ufl.edu/students/get-started/]. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

4.6 Diversity and Inclusion

The instructors of this course are committed to supporting all our students—to making class administration fair for everyone, to ensuring class space is welcoming for all, and to assignments and discussions as productive opportunities for all. We invite conversation with you about how we may better accomplish this at any time.

4.7 Course Evaluations

Students are expected to provide feedback on the quality of instruction in this course based on ten criteria. These evaluations are conducted online at: https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at https://evaluations.ufl.edu/results.

5. Resources

5.1 Technical support

If you have questions regarding Canvas or related technology used in connection with this course, you should contact the UF Computing Help Desk through one of the following:

Email: helpdesk@ufl.edu

Web: https://helpdesk.ufl.edu/
 Phone: (352) 392-HELP (4357)

One thing in particular that might cause some issues for you is seeing the marginal or "in-text" feedback on writing assignments. See Canvas's instructions [https://community.canvaslms.com/t5/Student-Guide/How-do-I-view-annotation-feedback-comments-from-my-instructor/ta-p/523] for seeing that feedback. If you have trouble, I can always individually prepare for you a PDF version of your work with comments easily visible, so let me know if you need me to do that.

5.2 Other support services

- U Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit the U Matter, We Care website [https://umatter.ufl.edu/] 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 [https://counseling.ufl.edu] 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 [https://shcc.ufl.edu/].
- University Police Department: Visit the UF Police Department website [https://police.ufl.edu/] 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 [https://ufhealth.org/emergency-room-trauma-center].
- Library Support [https://cms.uflib.ufl.edu/ask] provides various ways to receive assistance with respect to using the libraries or finding resources.

5.3 Perusall privacy and accessibility

You can find Perusall's privacy policy here and its accessibility statement here.

PHI 3681: Ethics, Data, and Technology (Dr. James Gillespie) Schedule of Readings and Assignments

This is the version of the schedule at the start of the term. The most up-to-date schedule will be found on the Canvas site.

Unit I: The Alignment Problem (4 Sections)

Week 1

- 01. Tue. 01/09 Introduction
 - No readings
- 02. Thurs. 01/11 The Alignment Problem
 - Christian, The Alignment Problem, Ch. 1 ("Introduction")

Week 2

- 03. Tue. 01/16 Do Artifacts Have Politics?
 - Winner, "Do Artifacts Have Politics?"
- 04. Thurs. 01/18 Al Safety
 - Amodei et al., "Concrete Problems in Al Safety"
 - USC Center for Artificial Intelligence in Society, "Social Networks and Substance Abuse Prevention for Homeless Youth" (webpage with videos)

Week 3

- 05. Tue. 01/23 Consequentialism I
 - Shafer-Landau, "Consequentialism: Its Nature and Attractions"
- 06. Thurs. 01/25 Consequentialism II
 - No new readings
 - Individual responses to case study 1 due 01/29

Week 4

- 07. Tue. 01/30 Moral Goals and Moral Constraints
 - Nozick, "Moral Constraints and Moral Goals"
 - Wells et al., "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show"
 - Lewis, "'Our minds can be hijacked': The Tech Insiders Who Fear a Smartphone Dystopia" (The Guardian)
 - Group responses to case study 1 due 01/31

Unit II: Autonomy (3 Sections)

Week 4

- 08. Thurs. 02/01 Digital Minimalism
 - Aylsworth and Castro, "Is There a Duty to Be a Digital Minimalist?"

Week 5

09. Tue. 02/06 Fake News and Free Speech

- Frenkel et al., "Surge of Virus Misinformation Stumps Facebook and Twitter" (The New York Times)
- Pontin, "The Case for Less Speech" (Wired)
- 10. Thurs. 02/08 Mill on Free Speech I
 - Mill, On Liberty, Ch. 2 ("Of the Liberty of Thought and Discussion") (excerpts)

Week 6

- 11. Tue. 02/13 Mill on Free Speech II
 - No new readings
- 12. Thurs. 02/15 Is the Attention Economy Noxious?
 - Castro and Pham, "Is the Attention Economy Noxious?"
 - Individual responses to case study 2 due 02/14

Unit III: Privacy (4 Sections)

Week 7

- 13. Tue. 02/20 Why Does Privacy Matter? I
 - Rachels, "Why Privacy is Important"
 - Amnesty International, "Surveillance Giants," pages 1-26
- 14. Thurs. 02/22 Why Does Privacy Matter? II
 - Marlinspike, "Why 'I Have Nothing to Hide' is the Wrong Way to Think About Surveillance"
 - Group responses to case study 2 due 02/21

Week 8

- 15. Tue. 02/26 The Traditional Privacy Protection Paradigm I
 - Barocas and Nissenbaum, "Big Data's End Run Around Anonymity and Consent"
- 16. Thurs. 02/29 The Traditional Privacy Protection Paradigm II
 - No new readings
 - Midterm exam due 03/01

Week 9

- 17. Tue. 03/05 Anonymization and Differential Privacy I
 - Kearns and Roth, The Ethical Algorithm, "Algorithmic Privacy: From Anonymity to Noise"
- 18. Thurs. 03/07 Anonymization and Differential Privacy II

Week 10

- 19. Tue. 03/12 Spring Break
- 20. Thurs. 03/14 Spring Break

Week 11

- 21. Tue. 03/19 A Puzzle about Privacy
 - No new readings
 - Individual responses to case study 3 due 03/18

Unit IV: Fairness (2 Sections)

Week 11

- 22. Thurs. 03/21 Algorithmic Discrimination I
 - Fazelpour and Danks, "Algorithmic Bias: Senses, Sources, Solutions"
 - Altman, "Discrimination," sections 1–2 (Stanford Encyclopedia of Philosophy)

Week 12

- 23. Tue. 03/26 Algorithmic Discrimination II
 - Angwin et al., "Machine Bias" (ProPublica)
 - Corbett-Davies et al., "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear." (Washington Post)
 - Group responses to case study 3 due 03/25
- 24. Thurs. 03/28 Algorithmic Discrimination III
 - Mayson, "Bias in, bias out," pp. 2221–2250

Week 13

- 25. Tue. 04/02 Algorithmic Discrimination IV
 - Mayson, "Bias in, bias out," pp. 2251–2261

Unit V: Transparency (3 Sections)

Week 13

- 26. Thurs. 04/04 The Black Box Problem
 - Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms"
 - Individual responses to case study 4 due 04/08

Week 14

- 27. Tue. 04/09 The Right to Explanation
 - Vredenburgh, "The Right to Explanation"
- 28. Thurs. 04/11 The Right to Explanation
 - No new readings
 - Group responses to case study 4 due 04/15

Week 15

- 29. Tue. 04/16 Transparency as Surveillance I
 - Nguyen, "Transparency Is Surveillance"
- 30. Thurs. 04/18 Transparency as Surveillance II
 - No new readings

Week 16

31. Tue. 04/23 Conclusion

FINAL EXAM due 04/28