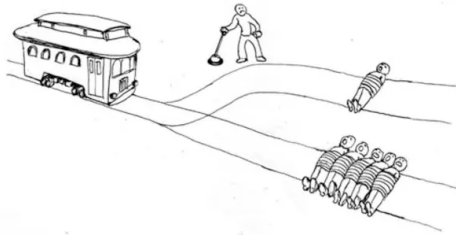


## INST 204: Designing Fair Systems



**Graduate TA:** Maria Isabel Magaña Arango  
([mimagar@umd.edu](mailto:mimagar@umd.edu))

**Office hours:** We/Th 10am–11am by appointment  
<https://calendly.com/marisamagar/15-minute-ta?> (Zoom or HBK Info Commons as preferred)

**Instructor:** Daniel Greene  
([dgreene1@umd.edu](mailto:dgreene1@umd.edu))

**Undergraduate TA:** Deena Habash  
([dhabash@terpmail.umd.edu](mailto:dhabash@terpmail.umd.edu))

**Office hours:** Tu/Th 3:30-4:30 PM by appointment:  
<https://calendly.com/dgreene1/office-hours>  
(Zoom or 4105HBK as preferred)

**Class time:** Tu/Th 2-3:15 PM ESJ 2212

"...the best way to bring about the ideal is by recognizing the nonideal, and that by assuming the ideal or the near-ideal, one is only guaranteeing the perpetuation of the nonideal."

- Charles W. Mills  
(2005)

### Catalog Description

Explores how policy shapes design and how design can act as de facto policy, focusing on emergent issues of algorithmic bias. Introduces students to interdisciplinary research on fairness, accountability, transparency, and justice in technical systems, bringing together fields such as law, computer science, and sociology. Students will learn how to assess the impact of automated decision-making in domains such as criminal justice and transportation, conduct audits of these systems, and communicate their results to the public and key decision makers.

### Course Description

This class reviews how automated decision making systems are designed, used, and regulated, and what political and ethical values enter the system at each stage. These values create differential outcomes for the different people enmeshed in these systems, but both these values

and these systems can be changed to support different values and different outcomes. In this way, the class serves as an introduction to the emerging field of *algorithmic bias* that bridges the disciplines of information science, computer science, law, policy, philosophy, sociology, urban planning, and others. Researchers and advocates in this field are currently leading investigations of the opaque, machine-learning driven systems that distribute rides, jobs, bail, loans and more in nonprofits, governments, and corporations. Following the leadership of such academic conferences as FAccT, this class is built around three specific values and corresponding case studies of automated decisionmaking: Fairness and policing, accountability and self-driving cars, and transparency and social media moderation.

We will spend one of our class sessions each week discussing technical and political issues in algorithmic bias, and the other in a 'lab' that applies those ideas to real-world examples. Readings bridge news coverage of algorithmic bias, philosophical consideration of the values in question, applied studies of these systems, artwork that communicates the danger to the public, and briefs from activists and wonks trying to change the system. Students will try out various means of auditing these systems and various methods of communicating their critiques to both the lay public and powerful decision makers. The differential impact of these systems across race, class, ability, and gender will be a priority throughout. Because these problems are new and of great importance, we will inevitably be pushed out of our comfort zones: Students strong in writing and reading will have to do more math than they're used to, while students strong in programming will be pushed to engage in heady philosophical debates without clear benchmarks for success.

## Student learning outcomes

At the conclusion of this class, you should be able to:

- Explain how social values influence technology design at the individual, organizational, and societal levels.
- Articulate the specific values built into automated decisionmaking systems.
- Identify the differential impact of these systems across social groups in a stratified society.
- Plan, design, and execute an audit of algorithmic bias.
- Communicate the consequences of algorithmic bias and the results of an audit to the lay public and key decision makers

## Acknowledgements

This class would not be possible without the inspiration of the following colleagues, in their research, teaching, and activism: Sarita Schoenbeck, Lisa Nakamura, Nick Seaver, Solon Barocas, Sarah Hamid, Anna Lauren Hoffman, Meredith Whittaker, and Alex Hanna.

## Assignments

*Each assignment will have a detailed rubric. Assignments add up to 1000 points. All assignments are due on Thursdays at midnight unless otherwise stated in the rubric.*

### **Participation (200 points, individual)**

Students are not assessed on attendance but rather on their active participation in class. Participation will look different for different students. It might mean asking clarifying questions, responding to instructor prompts, or constructively disagreeing with a peer. Different class sessions will also require different kinds of participation, from open discussion to group work on a data set. No matter what student or class, participation always requires active engagement with our readings before class and completion of an in-class activity: either an informal writing prompt (your 'exit ticket') at the start of class, or your signature on a group activity. Students have four 'freebie' days in which they can miss class for any reason with no penalty.

### **Reading Quizzes (200 points, individual)**

Quizzes will assess student engagement with the readings. They will be weekly, opening first thing Wednesday and closing midnight Friday. These are simple reading checks and do not require mastery of complicated concepts--that's what class discussions and assignments are for. Only the 10 highest-scoring quizzes will be graded, meaning students may choose to skip up to three quizzes with no penalty.

### **Facial recognition brief (100 points, individual)**

Students will compose an open letter to UMD President Darryl Pines arguing for or against the use of automated facial recognition on UMD's campus. This is our first assignment and serves as both an introduction to the ideas and methods of this course, and an early opportunity for instructors to assess students' writing.

### **Audit (200 points, group)**

Each of our three main units will introduce an applied case study of a particular automated decision-making system and the biases within it. For example, our unit on fairness and policing will include a class lab exploring the thousands of (anonymized) individuals enrolled in the Chicago Police Department's now-defunct gang database. The first two units, on fairness and accountability, will offer students the opportunity to pursue a more formal audit of that system and communicate the results to a wider audience. Students must complete one of these two audits as part of a pair or a group of three. Both audits require an outline and meeting with the instructor prior to submission. Students may choose to complete both audits in the same or different groups, but only the highest scoring audit will count towards your final grade.

### **Case Study (300 points, group)**

For our final assessment, student will explore an automated decision making system that is not covered by this class. In pairs or groups of three, students will define the inputs and outputs of this system and any biases that are visible in the process, before suggesting interventions to improve the system. The case study can be communicated in any form that is *not* a longform piece of writing. Example products could include a short film, a series of poems, a text-based game, or an annotated playlist. The assignment will be scaffolded such that multiple deliverables (e.g., research plan, outline) will add up to 300 points.

## Grade Evaluation

**Missed Deadlines:** If you think you will not be able to meet an assignment deadline for whatever reason, contact Dr. Greene before the due date to explain why you will need to submit the assignment late and what your plan is. If you cannot solve the problem within the bounds of the syllabus, we will work with you to solve it.

**Late Assignments Policy:** Unless prior permission has been granted, no late work is accepted. This policy is in place to ensure every student has their work returned to them in a timely fashion. Please prepare in advance so that you will not encounter technical difficulties that may prevent submission of a given assignment.

**Grading:** The primary purpose of a grade is to assess how well you know the concepts, techniques, and tools that are the focus of the class, providing you and the instructor with a signal that indicates whether a change is needed or you can keep doing what you're doing. Each assignment will be graded based on a rubric available to you and will be based on mastery of concepts. If you believe that a grade you received does not accurately reflect your knowledge and ability (either due to a grading error or a trivial misunderstanding on your part), you may raise the issue within one week of receiving the grade by speaking with Dr. Greene in office hours. After considering the issue, Dr. Greene will adjust your grade (either up or down) to best reflect your knowledge of the material.

**Final Grades:** Final grades will be submitted 48-72 hours after final exams are submitted. Because grades are issued in points, rather than percentages, there is no rounding up or down. Because grades are calculable throughout the semester, I will not respond to email requests for a grade bump at the end of the semester.

The cutoffs are as follows:

A+ 970-1000 points	A 930-969	A- 900-929
B+ 870-899	B 830-869	B- 800-829
C+ 770-799	C 730-769	C- 700-729

D+ 670-699	D 630-669	D- 600-629
F less than 600 points		

In this class, an "A" denotes full achievement of the goals of the class, a "B" denotes good progress towards the learning objectives, and a "C" indicates that you were able to comprehend the concepts involved but were unable to effectively apply that knowledge. Since the grading is based on a point-based system, an F is not the same thing as a zero. Failing work still earns some points. You are always better off to turn something in and get feedback on what you were able to complete. The point-based system also means that you can keep track of your progress and your current grade. You are encouraged to continuously monitor your own performance.

## Policy on Academic Misconduct

Cases of academic misconduct will be referred to the Office of Student Conduct irrespective of scope and circumstances, as required by UMD's rules and regulations. Instructors do not have a choice of following other courses of actions in handling these cases. There are severe consequences of academic misconduct, some of which are permanent and reflected on the student's transcript. For details about procedures governing such referrals and possible consequences for the student please visit <http://osc.umd.edu/OSC/Default.aspx>

It is very important that you complete your own assignments, and do not share any files or other work. The best course of action to take when a student is having problems with an assignment question is to contact the instructor or teaching assistant. They will be happy to work with students while they work on the assignments.

### **University of Maryland Code of Academic Integrity**

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit <http://shc.umd.edu/SHC/Default.aspx>

Students have a responsibility to familiarize themselves with violations of the Code of Academic Integrity. Among these include:

1. *Cheating* "Intentionally using or attempting to use unauthorized materials, information, or study aids in any academic exercise."
2. *Fabrication* "Intentional and unauthorized falsification or invention of any information or citation in an academic exercise."

3. *Facilitating Academic Dishonesty* "Intentionally or knowingly helping or attempting to help another to commit an act of academic dishonesty."
4. *Plagiarism* "Intentionally or knowingly representing the words or ideas of another as one's own in an academic exercise."

## UMD Resources for Students in Need

Students with disabilities should inform Dr. Greene of their needs at the beginning of the semester. Please also contact the Disability Support Services (301-314-7682) or <http://www.counseling.umd.edu/DSS/>. DSS will make arrangements to determine and implement appropriate academic accommodations.

Students who want help improving their writing are encouraged to visit UMD's Writing Center, where trained coaches will help you plan out assignments or edit drafts:  
<http://www.english.umd.edu/academics/writingcenter>

Students seeking more general support on specific skills necessary to be successful in college, or just looking for general help on how to manage their workload, are encouraged to visit Learning Assistance Services: <https://lasonline.umd.edu/>

Students encountering psychological problems that hamper their course work are referred to the Counseling Center (301-314-7651 or <http://www.counseling.umd.edu/> for expert help. For more information on UMD's Student Services, see <http://www.studentaffairs.umd.edu/student-life>

If you or someone you know has trouble procuring food please visit the campus food pantry:  
<http://campuspantry.umd.edu/>

If you or someone you know feels unsafe, the university has resources (see list below). Please note that both Dr. Greene and the TA are mandatory reporters under the UMD sexual misconduct policy, meaning that if we hear about sexual misconduct we are required by law to report it to the University for further investigation. Students impacted by sexual assault should contact OCRSM (below) for administrative support and CARE to Stop Violence for counseling support: <http://health.umd.edu/care>

Read more about hate-based crimes here:  
[https://ocrsm.umd.edu/files/Hate\\_Bias\\_FAQs\\_final.pdf](https://ocrsm.umd.edu/files/Hate_Bias_FAQs_final.pdf)

Counseling Center	301-314-7651 / <a href="http://www.counseling.umd.edu">www.counseling.umd.edu</a>
University Health Center and Mental Health Services	301-314-8180 / <a href="http://www.health.umd.edu">www.health.umd.edu</a>

University of Maryland Chaplains	thestamp.umd.edu/memorial_chapel/chaplain s
Office of Civil Rights and Sexual Misconduct	301-405-1142 / www.ocrsm.umd.edu

## Dr. Greene's Tips for a Successful Semester

1. Struggling with classes in general? Talk to the professor, the TA, friends, family, and/or the Counseling Center. We want to help you do well.
2. Have a question outside of class? After checking the syllabus and ELMS, feel free to contact Dr. Greene or the TA via email. Make sure to include the course number (INST208) in the subject line to ensure they see it. They will respond within 48 hours. Do not email multiple times if you have not received a response and less than 48 hours have passed. Send your emails to the appropriate party: Small paperwork questions to the TA, bigger conceptual or procedural questions Dr. Greene. Unsure what a professional email looks like? That's OK! Take a look at this guide: <https://medium.com/@lportwoodstacer/how-to-email-your-professor-without-being-annoying-af-cf64ae0e4087>
3. Visit us during office hours to talk about course content or anything else on your mind. In general, if you think your question or the answer you need for it is longer than a sentence or two, it's better to talk in person than over email.
4. Know your rights as an undergraduate student at UMD: University of Maryland Policies for Undergraduate Students <https://www.ugst.umd.edu/courserelatedpolicies.html>
5. Have fun! This class is designed to take on familiar problems in surprising ways. Embrace the weird and you'll do great.

## Course Schedule

### Unit 1: Introductions

#### **Week 1: Stakes and claims**

Tuesday, August 29

- Meet each other and the syllabus

Thursday, August 31

- Watch *Coded Bias* (2020) on Netflix or ELMS
- [Graham \(2012\) "A guide to reading and analyzing academic articles."](#)

## Week 2: How do machines learn?

Tuesday, September 5

- [Hanna, Alex, Emily Denton, Razvan Amironesei, Andrew Smart, and Hilary Nicole \(2020\). "Lines of sight." \*Logic Magazine\* 12.](#)
- Play with the [Datasets Have Worldviews](#) tutorial

Thursday, September 7

- [Garvie, Clare. \(2019\). \*Garbage in, garbage out: Face recognition on flawed data.\* Georgetown Law Center on Privacy and Technology.](#)
- [Agre, Phil \(2003\). "Your face is not a barcode."](#)

Unit 2: Fairness, prediction, and policing

## Week 3: How do you make a machine 'fair?'

Tuesday, September 12

- [Friedler, Sorelle, Carlos Scheidegger, and Suresh Venkatasubramanian \(2021\). "The \(im\)possibility of fairness: Different value systems require different mechanisms for fair decision making." \*Communications of the ACM\* 64\(4\): 136-143.](#)
- [Constantaras, Eva et al. 2023. "Inside the suspicion machine." \*Wired\*, March 6.](#)

Thursday, September 14

- [Read and watch "Long Island Divided"](#)
- [Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. \(2014\). Auditing algorithms: Research methods for detecting discrimination on internet platforms. \*Data and discrimination: converting critical concerns into productive inquiry\*, 22, 4349-4357.](#)

Friday September 15

- Facial Recognition Brief Due

## Week 4: Automating justice

Tuesday, September 19

- [Lum, Kristian and William Isaac \(2016\). "To predict and serve?" \*Significance\* 14 \(5\): 14-19.](#)



- [Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner \(2016\). "Machine bias." \*ProPublica\*.](#)

Thursday, September 21

- Smith, Larry (2018) ["Former Baltimore Police Detective Criticizes the Department's Gang Database."](#) *The Appeal*.
- Trujillo, Josmar & Alex Vitale (2019). ["Gang takedowns in the De Blasio era: The dangers of precision policing."](#)

### **Week 5: Who is risky?**

Tuesday, September 26

- Lavigne, Sam, Francis Tseng, and Brian Clifton (2017). [White paper](#) and [project](#) for "White collar crime risk zones." *The New Inquiry*.
- [Begely, Josh \(2016\). "Office involved." \[exhibit\]](#)

Thursday, September 28

- [Stop LAPD Spying Coalition \(2018\). "Before the bullet hits the body: Dismantling predictive policing in Los Angeles." \[Focus on 29-47\].](#)
- [Stop LAPD Spying Coalition \(2021\). "Automating banishment." \[Focus on parts 5, 6, and the Conclusion.\]](#)

### **Week 6: Fairness vs justice**

Tuesday, October 3

- [Keyes, Os, Jevan Huston, and Meredith Durbin \(2019\). "A mulching proposal." \*Proceedings of the ACM Conference on Human Factors in Computing\*.](#)

Thursday, October 5

- Wang, Kapoor, Barocas, and Narayan (2023). "Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy." FAccT.
  - Make sure to review all the little toggles in the [website](#) and the [rubric they developed](#)

Unit 3: Accountability and autonomous vehicles

### **Week 7: Can you sue a robo-car?**

Tuesday, October 10

- [Bogost, Ian \(2018\). "Can you sue a robo-car?" \*The Atlantic\*.](#)
- [Smiley, Lauren \(2022\). "'I'm the operator': The aftermath of a self-driving tragedy." \*Wired\*. \[PDF in case you've used up your free articles\]](#)
- [Play the MIT Moral Machine game.](#)

Thursday, October 12

- [Nissenbaum, Helen \(1996\). "Accountability in a computerized society." \*Science and Engineering Ethics\* 2: 25-42.](#)

Friday, October 13

- Audit 1 due

### **Week 8: How autonomous are they?**

Tuesday, October 17

- [Brandom, Russell \(2018\). "Self-driving cars are headed toward an AI roadblock." \*The Verge\*.](#)
- [The Daily \(2023\). "Ready or not, driverless cars are here." \*The New York Times\* \[podcast\]](#)

Thursday, October 19

- Asaro, Peter, 2011. "A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics," in P. Lin, K. Abney, and G. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge, MA: MIT Press. [PDF]
- Review the NHTSA self-driving car page <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>

### **Week 9: Regulating autonomy**

Tuesday, October 24

- Elish, Madeleine Clare (2019). "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." *Engaging Science, Technology, and Society* 5: 40-60.

Thursday, October 26

- Stilgoe, Jack. (2018). [We Need New Rules for Self-Driving Cars. \*Issues in Science And Technology\*, 34\(3\), 52-57.](#)

- Vellinga, Nynke E. (2017). "From the testing to the deployment of self-driving cars: Legal challenges to policymakers on the road ahead." *Computer Law & Security Review* 33 (6): 847-863.

## **Week 10: Autonomy and us**

Tuesday, October 31

- [Gregory, Derek \(2015\). "Little boys and blue skies: Drones through post-atomic eyes" \[lecture\]](#)

Thursday, November 2

- Suchman, Lucy. (2020). "Algorithmic warfare and the reinvention of accuracy." *Critical Studies on Security* 8(2): 175-187.
- Play [Unmanned](#) by Molleindustria

Unit 4: Transparency and social media

## **Week 11: Who, or what, are the mods?**

Tuesday, November 7

- [Seetharaman, Deepa, Jeff Horwitz, and Justin Scheck \(2021\). "Facebook says AI will clean up the platform. Its own engineers have doubts." \*The Wall Street Journal\*, October 17.](#)
- [Buni, Catherine and Soraya Chemaly \(2016\). "The secret rules of the internet." \*The Verge\*.](#)
- [Hopkins \(2017\). "Revealed: Facebook's internal rulebook on sex, terrorism and violence." \*The Guardian\*.](#)
- Review [Facebook's community standards](#)

Thursday, November 9

- Gillespie, Tarleton (2018). "Three imperfect solutions to the problem of scale" in *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press: 74-110. [PDF]
- 

Friday, November 10

- Audit 2 Due

## Week 12: Other possible platforms

Tuesday, November 14

- [Salehi \(2021\). "Do No Harm." Logic Magazine.](#)
- Explore [the moderation policies of fanfileIntroFour c site AO3](#), compare with Facebook's.

Thursday, November 16

- [Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Iii, H.D. and Crawford, K., 2021. Datasheets for datasets. \*Communications of the ACM\*, 64\(12\), pp.86-92.](#)
- [Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model cards for model reporting." In \*Proceedings of the conference on fairness, accountability, and transparency\*, pp. 220-229.](#)

## Week 13: Thanksgiving

Tuesday, November 21

- No class Tuesday or Thursday for Thanksgiving Break. Instructors will be available for office hours Tuesday.

Unit 5: Abolition

## Week 14

Tuesday, November 28

- Ananny, Michael. and Kate Crawford (2018). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *new media & society*, 20(3), pp.973-989. [PDF]

Thursday November 30

- [Benjamin, Ruha \(2020\). "The tiny, high-tech wolf in sheep's clothing." \*LEVEL\*.](#)
- Mackenzie, Donald. (1996). "Uninventing the bomb?" *Medicine, Conflict and Survival* 12(3): 202-211. [PDF]

## **Week 15**

Tuesday, December 5

- Workshop final projects

Thursday, December 7

- Workshop final projects

Monday, December 11

- Final Projects Due