# Georgia Institute of Technology
# Course Syllabus: CS6603 AI, Ethics, and Society

## Summer 2023
Delivery: 100% Web-Based on Canvas
Dates course will run: May 15, 2023 – July 31, 2023

## Instructor
Dr. Mahender Mandala
Email: mandala@gatech.edu

## Course Creator
Dr. Ayanna Howard
Email: ah260@gatech.edu

## Head TA
Vijay Sivanesan, P.E., PMP
Email: vsivanesan3@gatech.edu

## General Course Information

**Description**

Abuse of big data means your worst fears can come true. Are they being monitored by your employer? Check. Government intrusions into your daily life? Check. Being turned down by college admissions because you are predicted not to donate in 10-20 years? Check. Sounds a bit like the visions in the Minority Report. Alas, machine learning algorithms are already being deployed by industry, government, and, yes, even schools to make decisions that impact us in direct ways. Such programs are typically promoted as fair and free of human biases, but humans who make mistakes are programming, calibrating, and evaluating their performance. Thus resides the problem. How do we design algorithms that effectively deal with large amounts of data to train them while ensuring their outcomes aren't misused? In this course, not only will we examine various AI/ML techniques that can be used to counterbalance the potential abuse and misuse of learning from big data, but we will focus on the effects of these technologies on individuals, organizations, and society, paying close attention to what our responsibilities are as computing professionals.

**Pre-Requisites**

None. Although, it is expected that, throughout the course, you will develop working knowledge of using Python (which will help you complete the assignments later in the semester). Several optional exercises and additional lecture modules are provided to help you become familiar with Python and using Jupyter notebooks.

**Course Goals and Learning Outcomes**

There are several outcomes for the course, based on four primary modules:

*Module 1 - Data, Individuals, and Society*

Objective: After completing this module, students will be able to understand the power and impact that analytics and AI/ML have on individuals and society, especially concerning issues such as fairness and bias, ethics, legality, data collection, and public use.

*Module 2 – The BS of Big Data*

Objective: After completing this module, students will be able to understand the underlying components of big data, apply basic statistical techniques to data scenarios, and understand the issues faced when learning from big data, ranging from data biases, overfitting, causation vs. correlation, etc.

*Module 3 – Fairness in AI/ML*

Objective: After completing this module, students can understand and apply basic AI/ML techniques to data scenarios, focusing on identifying fairness and bias issues in designing decision-making systems. We will work systematically towards understanding technical approaches to current AI/ML applications, such as facial recognition, natural language processing, and predictive algorithms, all while being mindful of its social and legal context.

*Module 4 – Bias Mitigation and Future Opportunities*

Objective: After completing this module, students can utilize tools and methods to quantify bias and examine ways to use algorithmic fairness to mitigate it, considering ethical and legal issues. Students will apply their knowledge of analytics and AI/ML to transform a current biased dataset into a more objective solution.

In this class, you will be challenged to broaden your understanding of state-of-the-art AI/ML algorithms and solutions, considering their potential impacts on society. You will have ample opportunity to analyze various situations critically and viewpoints provided in papers, books, on the web, and from your observations. You will be able to practice your learned knowledge by writing coherent and healthy-structured critiques of situations and papers, leading and participating in class discussions, and designing your algorithmic solutions. The issue of data misuse and abuse is not easily solvable; concrete right or wrong answers are not easily determined until solutions are typically deployed into society. Given this, you are entitled to your opinions on any topics presented throughout the course, whatever they happen to be. You will not be penalized for your viewpoints; however, you must be able to support your viewpoints and resulting solutions effectively. This means showing that you have given your approach to a problem some thought, can discuss its various trade-offs and implications and can support other viewpoints, even though your personal views may differ.

## Course Materials

**Course Text**

*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* by Cathy O'Neil (2016)

**Additional Materials/Resources**

Additional assigned readings will be included with each assignment.

**Classroom Management Tools**

- Video Lectures: All video lectures are located on Canvas.
- Projects: are located on Canvas.
- Graded Discussions: are located on Canvas.
- Reading Materials: are located on Canvas.
- **Ed** Discussion: is located on Canvas.
- Grades: is located on Canvas.
- Exams: are located on Canvas using Honorlock

## Course Requirements, Assignments & Grading

**Assignment Distribution and Grading Scale**

| Assignments | Weight |
|---|---|
| Homework Projects | 40% |
| Written Critiques | 10% |
| Mid-Term Exam | 10% |
| Final Project | 15% |
| Final Exam | 10% |
| Class Discussion/Exercises *(Case Studies and, Exercises)* | 15% |
| **Total** | **100%** |

**Grading Scale**

Your final grade will be assigned as a letter grade according to the following scale:

A 90-100%

B 80-89%

C 70-79%

D 60-69%

F 0-59%

**Assignments Due Dates (Time zone)**

Assignment due dates are Sundays at 11:59 PM Anywhere on Earth time. We will only accept assignments submitted on time due to time zone issues. We recommend changing your time zone in Canvas to show the due date in your local time. There are no exceptions.

**Assignment Formats for Submittal**

All reports for assignments, projects and written critiques will be submitted as PDF in JDF Format. The JDF format is available in Latex, Google Docs, Microsoft Word and Pages document templates. There are no exceptions for assignments that do not use JDF format.

**Late and Make-up Work Policy**

Homework Projects and Written Critique assignments will be accepted with a deduction of 10% per 24-hour period starting after the due date submission time. Assignments over 3 days late (i.e., three 24-hour periods) will not be accepted. No time extensions are provided for the Exams, Final Project, and Class discussions/Exercises. There will be no make-up work provided for missed assignments. ***Please verify the information in your assignment before submission. We will NOT be able to make exceptions for submitting the wrong assignment or insufficient or otherwise inadequate information in the assignment.***

Of course, emergencies (illness, family emergencies) will happen. In those instances, please get in touch with the Dean. The Dean of Students is equipped to verify emergencies and pass confirmation on to all your classes. For consistency, we ask all students to do this in an emergency.

**Regrade Policy**

If you consider that your score is incorrect based on the feedback provided, you may request a rescore during the regrade time period, which is 7 full days from the date the grades are released. If the grades are released Monday, you will have till next Monday's End-Of-Day.
This request is only valid if it includes a valid and proper explanation of where the TA-grader made an error.
The student should open a private Ed post and copy all instructors and provide the following (per question to regrade) in the specified format:

- Question to regrade:
- Why do you consider there is an error in the score?

Remember that any rescore request may trigger a rescore of the entire assignment, resulting in other points being deducted.

Only one (1) rescore regrade per assignment is allowed, so ensure you provide enough evidence in your request. Please double-check before submitting, as we will consider only the first request and ignore the rest.

**Office Hours**

This class uses the chat tool, Ed for its office hours. **Ed** office hours are not scheduled at specific times; instead, the teaching team will typically respond quickly to **Ed** by end-of-day. You can also email or post privately on **Ed** or set up a chat via an alternate technology.

**Feedback**

Every semester, we make changes and tweaks to the course formula. As a result, we try new things every semester, and some of these things may need to be revised. We ask for your patience and support as we figure things out, and in return, we promise that we, too, will be fair and understanding, especially with anything that might impact your grade or performance in the class. Second, we want to consistently get feedback on improving and expanding the course for future iterations. You can take advantage of the feedback box on **Ed** especially if you want to gather input from others in the class), give us feedback on the surveys, or contact us directly via private **Ed** messages.

## Technology Requirements and Skills

**Computer Hardware and Software**

- High-speed Internet connection
- Laptop or desktop computer with a minimum of a 2 GHz processor and 2 GB of RAM
- Windows for PCs OR Mac iOS for Apple computers.
- Complete Microsoft Office Suite or comparable and the ability to use Adobe PDF software (install, download, open, and convert)
- Mozilla Firefox, Chrome, and Safari browsers (*Note: Honorlock requires students to use Chrome*)

**Canvas**

This class will use Canvas to deliver course materials to online students. ALL course materials and assessments will take place on this platform.

**Proctoring Information**

The midterm exam will be proctored with open notes, pen/paper, and a calculator. A proctored exam is similar to the one you would take in the classroom. These exams are delivered via a tool called Honorlock. Honorlock is an online proctoring service that allows you to take your exam from the comfort of your home. You DO NOT need to create an account, download software, or schedule an appointment in advance. Honorlock is available 24/7; a computer, a working webcam, and a stable Internet connection are needed. You will need Google Chrome and download the Honorlock Chrome Extension to start. You can download the extension at: www.honorlock.com/extension/install.

When you are ready to take the exam, you would log into CANVAS, go to the course, and click on the exam link. Clicking Launch Proctoring will begin the Honorlock authentication process, where you will take a picture of yourself, show your ID, and complete a scan of your room. Honorlock will be recording your exam session via the webcam and your screen. Honorlock also has an integrity algorithm that can detect search-engine use, so please do not attempt to search for answers, even if

it's on a secondary device. If you encounter any issues, you may contact them by live chat, phone (844-243-2500), and email (support@honorlock.com).

## Course Policies, Expectations & Guidelines

**Communication Policy**

You are responsible for knowing the following information:

1. Anything posted to this syllabus
2. Anything emailed directly to you by the teaching team (including announcements Or Posts via **Ed**) 24 hours after receiving such an email.

Because **Ed** announcements are emailed to you, you should check your Georgia Tech email once every 24 hours to remain updated on new information during the semester. Georgia Tech generally recommends that students check their Georgia Tech email once every 24 hours.

We generally prefer to handle communication via **Ed** to help with collaboration among the teaching team, but we understand **Ed** is not ideal for having information “pushed” to you. We may contact you via a private **Ed** post instead of an email. Still, if we do so, we will send email notifications immediately, bypassing your settings, to ensure you’re alerted. This type of communication will also spring under #2 above.

Note that this means you won’t be responsible for knowing information communicated in several other methods we’ll be using. You aren’t responsible for knowing anything posted to **Ed** that isn’t linked to an official announcement. You don’t need to worry about missing critical information so long as you keep up with your email and understand the documents on this website. This also applies reverse: we do not monitor our Canvas message boxes and may not respond to direct emails. If you need to get in touch with the course staff, please post privately to **Ed** (either to all Instructors or to an instructor individually) or tag the instructor in the relevant post.

**Online Student Conduct and (N)etiquette**

Communicating appropriately in the online classroom can be challenging. To minimize this challenge, it is important to remember several points of “internet etiquette.”

That will smooth communication for both students and instructors:

- *Read first, Write later*. Read the ENTIRE set of posts/comments on a discussion board before posting your reply to prevent repeating commentary or asking questions that have already been answered.
- *Avoid language that may come across as intense or offensive*. Language can be easily misinterpreted in written and electronic communication. Review email and discussion board posts BEFORE submitting. Humor and sarcasm may be easily misinterpreted by your reader(s). Try to be as matter-of-fact and professional as possible.

- *Follow the language rules of the Internet.* Do not write using all capital letters because it will appear as shouting. Also, the use of emoticons can be helpful when used to convey nonverbal feelings. ☺
- *Consider the privacy of others.* Ask permission before giving out a classmate's email address or other information.
- *Keep attachments small*. If it is necessary to send pictures, change the size to an acceptable 250kb or less (one free, web-based tool to try is picresize.com).
- *No inappropriate material.* Do not forward virus warnings, chain letters, jokes, etc., to classmates or instructors. The sharing of pornographic material is forbidden.

*NOTE: The instructor reserves the right to remove posts that are not collegial and do not meet the Online Student Conduct and Etiquette guidelines.*

**University Use of Electronic Email**

A university-assigned student e-mail account is the official university means of communication with all students at the Georgia Institute of Technology. Students are responsible for all information sent via their university-assigned e-mail account. Students who forward information to their university e-mail account are responsible for all information sent to any other e-mail account, including attachments. To stay current with university information, students must check their official university e-mail accounts and other electronic communications frequently and consistently. Recognizing that some communications may be time-critical, the university recommends that electronic communications be checked minimally twice a week.

**Plagiarism & Academic Integrity**

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. All students enrolled at Georgia Tech and all its campuses are to perform their academic work according to standards set by faculty members, departments, schools, and colleges of the university; cheating and plagiarism constitute fraudulent misrepresentation for which no credit can be given and for which appropriate sanctions are warranted and will be applied. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Any student suspected of cheating or plagiarizing on an exam, exercise, or assignment will be reported to the Office of Student Integrity, investigate the incident, and identify the appropriate penalty for violations.

Reusing your own work without acknowledgment is considered self-plagiarism. If you intend to use your work from a previous semester, please send us a private post in Ed, and during the assignment submission, point to the Ed post in the comments section in Canvas.

**All graphs shall be generated using Python, Excel, or utilizing other software tools, and Hand-drawn graphs, tables, and charts are not allowed.**

**Accommodations for Students with Disabilities**

If you are a student with learning needs that require unique accommodation, contact the Office of Disability Services at (404) 894-2563 or http://disabilityservices.gatech.edu/ as soon as possible to make an appointment to discuss your unique needs and to obtain an accommodations letter. Please also e-mail me as soon as possible to set up a time to discuss your learning needs.

**Student-Faculty Expectations Agreement**

At Georgia Tech, it is essential to strive for an atmosphere of mutual respect, acknowledgment, and responsibility between faculty members and the student body. See http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectations that you can have of me and that I have of you. Ultimately, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

**Subject to Change Statement**

The syllabus and course schedule may be subject to change. Changes will be communicated via the **Ed** announcement tool. Students must check **Ed for instructor posts,** email messages, and course announcements to stay current in their online courses.

# Course Schedule

| Week/Dates | Topics | Deliverables |
|---|---|---|
| **1**<br>**May 15** | **Lesson 1 Data Individuals, and Society Introduction**<br>**Lesson 2 Overview** | |
| **2**<br>**May 22** | **Lesson 3 Ethics vs. Law**<br>**Lesson 4 Data Collection** | |
| **3**<br>**May 29** | **Lesson 5 Fairness and Bias** | **Assignment: FB**<br>**Case Study: Emails Exposed**<br>**Case Study: Loan Denied** |
| **4**<br>**June 5** | **Lesson 6 BS of Big Data & Stats 101 Overview**<br>**Lesson 7 Python and Stats 101** | **Case Study: Facebook Manipulation**<br>**Assignment: Stats 101** |
| **5**<br>**June 12** | **Lesson 8 Descriptive Statistics**<br>**Lesson 9 Inferential Statistics: Sampling Bias**<br>**Lesson 10 Inferential Statistics: Causation vs. Correlation** | **Exercise: Anscombe's Quartet**<br>**Exercise: Smoking Sampling Bias Design**<br>**Written Critique: Ethical Autonomous Vehicles** |

| Week/Dates | Topics | Deliverables |
|---|---|---|
| 6<br>June 19 | Lesson 11 Inferential Statistics: Confidence<br>Lesson 12 AI/ML Techniques: Word Embeddings<br>Lesson 13 Bias in Word Embeddings | Mid-Term Exam (Ethics and Stats)<br>Case Study: Word Embeddings<br>Exercise: Bias in Word Embeddings |
| 7<br>June 26 | Lesson 14 AI/ML Techniques: Facial Recognition<br>Lesson 15 Bias in Facial Recognition | Assignment: AI/ML Part I<br>Case Study: Facial Recognition |
| 8<br>July 3 | Lesson 16 AI/ML Techniques: Predictive Algorithms<br>Lesson 17 Crime-based Predictive Algorithms | Assignment: AI/ML Part II<br>Exercise: Predictive Algorithms<br>Case Study: Predictive Algorithms |
| 9<br>July 10 | Lesson 18 Bias in Predictive Algorithms<br>Lesson 19 Fairness and Bias<br>Lesson 20 Fairness and Bias Assessment Tools | Written Critique: What-If Tool<br>Exercise: AI Fairness 360<br>Exercise: What-If Tool |
| 10<br>July 17 | Lesson 21 AI/ML Techniques for Bias Mitigation | Assignment: Fairness and Bias |
| 11<br>July 24 | Lesson 22 AI, Society, and Ethics Wrap-up | Final Project |
| 12<br>July 31 | | Final Exam |
| 13<br>August 3 | | End-of-Course Survey, CIOS Survey |