Workshop Proposal

Workshop Presenters: Theo, Claire, Dakota
Underclassmen Advisors: Nabilah, McKenzie
Workshop Title: System Error: Racism in AI
Faculty Advisors: Steve Russell, Kiran Bhardwaj

Working description of MLK Day workshop: (DRAFT)

Few realize the ubiquity and influence of AI algorithms in our lives. This workshop will 1) address the fact that AI's influence is invisible to most of us; 2) show how AI is biased against women and people of color in largely invisible ways, and 3) introduce participants to the great advocacy out there on the legislative and grassroots level seeking to address racial and gender bias

Possibility of an external virtual speaker?
If possible, Dr. Bhardwaj has said that she'd like to enlist an external scholar to virtually present or assist with the design of the workshop.

Notes w/ Ms(). O:

- Start with asking participants to call out the top 5 apps they've used most in past week
    - Or have people shout out all the places they think AI exists, and then rattle off a million more?
- Start with what is AI and where innovation is happening
- Start with misconceptions
- Framed it with 3 questions -- good
- How long is it -> too much content
    - Tying it more closely to MLK
    - Talk about policy and legislation (change to *be more specific*)
        - What is missing is existing platforms
    - Talking about conflicts that have happened already
        - Storytelling? -> "what is ai" and who is building it
        - OpenAI and google/ethics
    - Ms. O will send along her paper on MLK's potential response to 5 pressing topics (I have a digital dream)
        - Weaponizing AI
        - Lack of diversity in tech itself
- Don't just deliver to listeners but create active engagement and genuine change to other people not present
- Big tech legislation/breaking up big tech?
    - Privacy laws for big tech
    - GDPR in Europe set a good example
    - Social media / ad recommendations
- **"Move fast and break things"**

        - Mindset of the leadership
        - Value of productivity -> how to interrogate "work-forward" mindsets
            - There's a lot of focus on what we make per capita as opposed to spending time thinking about the ethics / effects
        - Game on hiring biases in AI: https://www.survivalofthebestfit.com/
- Make it clear that this is an area that *needs* activism in the first place
    - It's like a black box - don't know enough about it to talk about it
- **Deprioritize**:
    - Extend one topic across the three questions as an entry point/throughline
        - Include stories of current advocacy?
        - Approach this topics selection through the lens of "dismantling misconceptions"
    - Data and policing story -> carry through the entire workshop
    -

Detailed outline of your workshop content, goals, and outcomes: (Claire and Theo)

Our workshop is structured in a three-pronged approach in an attempt to address the fundamental issues at play: why AI is important, how AI is racist (or, perhaps better to say, how does it perpetuate biases), and how do we create non-biased AI/what does the future of AI looks like. Each of these will be handled by one of our student presenters, and they will be divided into the same structure: a quick presentation, some sort of demo, and then a facilitated conversation.

Why is AI important?

- **AI is ubiquitous.** This section of our workshop will demonstrate how AI is being used already in the real world and has effects on people's lives.
- Talk about (some of) the different subfields of AI and try to establish some categories – i.e. generating models, natural language processing, etc. – to differentiate between types of AI.
    - Could demo some AIs and ask the audience to classify them?
- AI is framed as having the potential to do good – AI could theoretically be used to *reduce* bias and noise.
    - Human error (failures of consistency; human bias), e.g. judicial reform and the failure of mandatory minimum sentencing

How is AI Racist?

- A brief explainer of how AI works – i.e. the idea of fitting to data and what happens when data contains not-so-good stuff (e.g. negative connotations with certain races, the propagation of vitriolic rhetoric, a 'preference' for certain kinds of faces, and much much more).
- Run through some of the most high-profile cases of racist AI, through different categories – facial recognition, natural language processing, etc.

    - Have the audience try it out themselves. We can test the bias of NLP or facial recognition models pretty easily whether by setting up our own demos or using something like https://hownormalami.eu
- Talk about some of the sketchy development.
    - The shadiness of OpenAI, the brute-force methods favored by AI companies.

What does it mean that AI is racist/how do we combat it/what is the future?

- Building our future technology by enshrining and amplifying societal biases will only further them.
    - Introduce work by the Algorithmic Justice League, Timnit Gebru's new AI ethics center, and other advocates.
- Talk about the problem of representation in the field of computer science and what it takes to change that.
    - Countless articles and projects – an example, and another talking about how qualified BIPOC candidates *do* exist and just aren't being hired.
    - Identify the problem, work on generating solutions, refer to https://qz.com/1945141/5-ways-to-get-more-us-students-of-color-into-computer-science/ or other resources with recommendations.
- Legislation, for example, talks about Biden's AI bill of rights and the EU's attempts to curtail privacy issues with AI.
-

Intended audience (introductory, intermediate, advanced):

- Introductory—no technical knowledge on AI or racial justice issues necessary.

Main takeaways of your workshop:

- Metacommentary: explain why AI and racial justice in AI are important, and the self-perpetuating nature of the problem.
    - Bias must be addressed and considered as AI is incorporated in more aspects of our lives because of its great risk of unknowingly exacerbating prejudice/discrimination against minorities
- Noise in human decision-making can be prevented by machines; how do you do it right?
    - Minimum sentencing failure in reducing judicial noise, furthering and enshrining bias
- Data is not perfect. Data does contain bias. There's no such thing as quantitative perfection.
    - Data collection is inherently biased both in terms of class (easier to access people of higher class), race (distrust and regions),
    - Dismantle techno-chauvinism.
- Actionable steps: invite students to consider racial bias in their future projects.
    - Important to have diverse teams, those developing AI models ought to reflect its consumers/users
    - Data ought to be transparent and ethically sourced.

How does the workshop connect to the legacy of Dr. King and the Civil Rights Movement: (Dakota and McKenzie)

At the annual MLK Now event, Alexandria Ocasio-Cortez said, “Algorithms are still made by human beings, and those algorithms are still pegged to basic human assumptions… They’re just automated assumptions. And if you don’t fix the bias, then you are just automating the bias,” (Vox). Dr. King’s legacy teaches us to fight for what is right. This workshop will introduce to students what A.I. is, how it is biased, and the adverse effects on people of color. By educating Andover students about the dangerous biases in A.I., this workshop will start a conversation and motivate students to rally for systematic change in the field of Artificial Intelligence.

Artificial intelligence (largely dependent on algorithms) has historically been developed by white males. Although computer science was a field jump-started by women, the more en vogue it’s become, the less diverse it has grown. Non-asian and non-white software engineers account for less than 2% of all software engineers, despite or really because of the fact that qualified BIPOC candidates exist and are left unhired. This small group of homogenous men is in control of software that will continue to grow in usage by the general population. It is dangerous and unjust to have an unrepresentative group making decisions that influence millions of lives – by having a diverse group of people leading the development of Artificial Intelligence, critical blindspots are avoided. Historically, there have been very few people able to question unethical or non-inclusive decisions made by this dominant group. Small steps have been taken to increase diversity in Artificial Intelligence, but an overwhelming campaign similar to the Civil Rights Movement must be ignited to create any actual change in the Artificial Intelligence community.

In addition, AI tools themselves must be ethically designed and created. Current models are often trained on indiscriminate webscrapes of large swaths of the internet – that’s an issue. Proliferation of bias in unscreened, unaudited data leads to inherently flawed models (models which can never be “fixed” because biased assumptions are baked into their very core). Furthermore, AI researchers often treat their models as big black boxes, unaccountable for the decisions they make. Transparent design – and transparent AI research – is paramount to any fight against further enshrining systemic discrimination.

Civil rights activism is a proactive effort, not just a reactive one. To work toward a less racist future, we have to act now, not tomorrow, to combat ubiquitous technological inequality. This workshop will allow us to educate on the breadth of the problems, their importance, and the steps students and those of the younger generation need to be prepared to take.

Structure:

- Presentation:
    - Why AI is important

Any special requests (room preference, number of students in the group, technology requirements?):

- Data lab within the Makerspace
-

To begin brainstorming, here's the general call for MLK day proposals and the link to submit a proposal (which includes a brief description, outline, and workshop goals). You decide what level of audience (introductory to more advanced), and size, etc.

MLK Day Description: On Monday, January 17th, Phillips Academy will celebrate the life and legacy of Martin Luther King, Jr. In the morning we will welcome a special guest speaker, Eddie Glaude, Jr., during All School Meeting and in the afternoon Uppers, Seniors and PostGraduates will participate in workshops led by their peers.

In keeping with our mission to provide opportunities for students to explore and discuss issues relating to diversity and multiculturalism and the school's commitment to equity and inclusion, we welcome workshops that explore issues including but not limited to race, class, gender and sexual identity, ability, ethnicity, intersectionality, etc.

Please note that workshops run between 60-75 minutes in length and should include some time for content sharing and discussion. All workshops MUST be grounded in historical context and presenters must work with a faculty advisor in developing and executing their workshop.

We invite you to submit a proposal for a workshop that explores issues including but not limited to race, class, gender and sexual identity, ability, ethnicity, intersectionality, etc.

The deadline to submit the proposal is Dec 5th, but the actual workshop details can be fleshed out a lot more after we return to campus in January.

Potential student co-facilitators:
Theo Baker
Claire Wang
Dakota Chang
Nabilah Nazar
Huda Abdulrasool
McKenzie Williams
Nadia Choophungart
Constantine Krenteras

Let's pool our thoughts below!
I think this workshop would be super exciting since the power of AI algorithms on our lives is 1) almost invisible to most of us and 2) is also biased against women and people of color in largely invisible ways, and because there's some great advocacy out there on the legislative and grassroots level to address these...

See Theo's op-ed on the topic HERE
The book "Weapons of Math Destruction" has a lot of excellent examples about misuse of AI algorithms.

Attached some useful sources below, worth taking a look at. Live demos are very engaging and it might not be a bad idea to try to incorporate them. Probably also worth having some firsthand accounts. Also, CS as a field is heavily discriminatory – we can't separate the tools from their development so this is absolutely something to talk about too. Would love to chat more anytime!
https://services.google.com/fh/files/misc/diversity-gaps-in-computer-science-report.pdf
https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities/
https://searchenterpriseai.techtarget.com/feature/Biden-sets-stage-for-national-AI-strategy
https://hownormalami.eu (awesome interactive website developed by EU demonstrating face detection algorithms; could be used as a jumping off point for our own demos if we want)
https://www.nature.com/articles/d41586-019-03228-6 (racial bias in health care algorithms)

And last thing: maybe worth talking about Noise and understanding the role AI and ML could have in *mitigating* bias and unfair treatment. I love **Kahneman** and a lot of his newest work has to do with statistical and ML models that can help reduce noisy systems (and Judea Pearl!).

There's a lot of health/medicine related data biases for AI, but maybe like hidden areas the bias in AI affects that people don't really think about? I think that'd be really interesting to talk about, such as hiring, policing & criminology, ads, banking (AI involved in lending money and other decisions?), NLP (a lot of gpt4 and gpt3 has hidden bias and inability to do sentiment analysis that people don't really think about) -- these are things I guess people might not think as much abt? But it might also be a step up from human biases (super cool article: Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms) - claire

Hi! I think we can also talk about the history of the industry and many popular but problematic datasets—for example, the Boston housing dataset or the 80 Million Tiny Images dataset (https://gizmodo.com/mit-takes-down-popular-ai-dataset-due-to-racist-misogy-1844244206). Giving these concrete examples might help bring the severity into context.
We can also mention more famous examples to relate them to our lives. For instance, Sam from Samsung could be a good representation of how human-made decisions could reinforce stereotypes. While she is not official, years of a feminine voice assistant reinforced stereotypes of women as assistants and allowed people to sexualize her. On a more technical aspect, we can also analyze the case of Tay and how techniques from NLP do not understand the context of words and can quickly become racist. However, it might be too complicated as we would mention stuff such as stemming and lemmatization.

> "Algorithms are still made by human beings, and those algorithms are still pegged to basic human assumptions," she told writer Ta-Nehisi Coates at the annual MLK Now event. "They're just automated assumptions. And if you don't fix the bias, then you are just automating the bias." *-- AOC (Alexandria Ocasio-Cortez says AI can be biased. She's right.)*

Another thing that would be cool to mention would be the echo chamber. A.I. club did a collaboration with PA Dems on this issue so I can put the outline of it in here as a resource if it helps. -- Dakota

An interesting example we could explore for hidden biases could be in the recruiting software that Amazon had to 'scrap' in 2018. The software relied on data from Amazon's current employees and it was used to highlight

certain candidates for potential jobs. However, since Amazon's employes were (and still are) mostly male, the AI was highlighting more male than female candidates.
(https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e)

For racial biases, we could dive into facial recognition software. There are tons of examples of facial recognition software failing to correctly identify people of color (i.e. Lamaya Robinson, Nijeer Parks, Robin Williams).

For context, Lamaya Robinson was denied entry to a roller rink because the rink's facial recognition falsely identified her as a person that was in a fight on the premises earlier. Nijeer Parks was "accused of shoplifting candy and trying to hit a police officer with a car... even though he was 30 miles away at the time of the incident". Robin Williams was falsely arrested after being accused of stealing watches and $3,800 worth of merchandise from a retail store. Facial recognition used by the Detroit Police Department accused Williams of the theft, despite the fact that Williams looked nothing like the actual thief.

These examples would cement the idea that the building of Artificial Intelligence is incredibly important. We could lean into the influence of who is in the 'room where it happens', so to speak.

Lamaya Robinson:
https://meaww.com/lamya-robinson-black-teen-kicked-out-skating-rink-after-facial-recognition-tech-misidentified-her

Nijeer Parks:
https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

Robin Wiliams:
https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig

- McKenzie Williams

One form of racial and gender bias in algorithms (briefly touched on by Claire earlier) is banking algorithms and specifically credit scores assigned, which are based off of prejudiced datasets. For example, using an Apple Card, one woman with a similar financial history to her husband received a credit line roughly twenty times lower than that of her husband's. Similar narratives have been shared by other women, where their gender has affected their credit card limits or credit scores at similar rates (in comparison to their husbands').

This shows a real-life situation that is incredibly important (because it's related to finances) that can affect women and other people in which their trustworthiness is wrongly decided due to biased data.

https://www2.deloitte.com/uk/en/pages/financial-services/articles/banking-on-the-bots-unintended-bias-in-ai.html

- Nadia Choophungart

I think that something else that we can touch on is how AI perpetuates economic disparities for marginalized groups. Bias in data sets used to train for the AI used for criminal legal systems, housing, workplace, and financial systems all perpetuate systemic biases that affect economic disparities in society.
https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities/

- Valerie

Another example of racial bias in AI is in self-driving cars. A Georgia tech study found that the machine vision used in self-driving cars was 10% less likely to identify black people as pedestrians than white people. Having racial diversity in the dataset was something the developers did not consider. While other radars/sensors that don't rely on visible light are sometimes used to prevent accidents, there is significant market push towards camera-only systems.

https://arxiv.org/pdf/1902.11097.pdf
https://www.theguardian.com/technology/shortcuts/2019/mar/13/driverless-cars-racist

- Constantine