Inclusive Futures: Ethical Implications of AI and Its Impact on Marginalized Communities

Ethics Roundtable: Association for Practical and Professional Ethics
March 20, 2024

CHONG HO ALEX YU, PH.D., D. PHIL.

PROFESSOR AND PROGRAM DIRECTOR OF DATA SCIENCE

HAWAII PACIFIC UNIVERSITY

Bias against minority groups is not a new issue

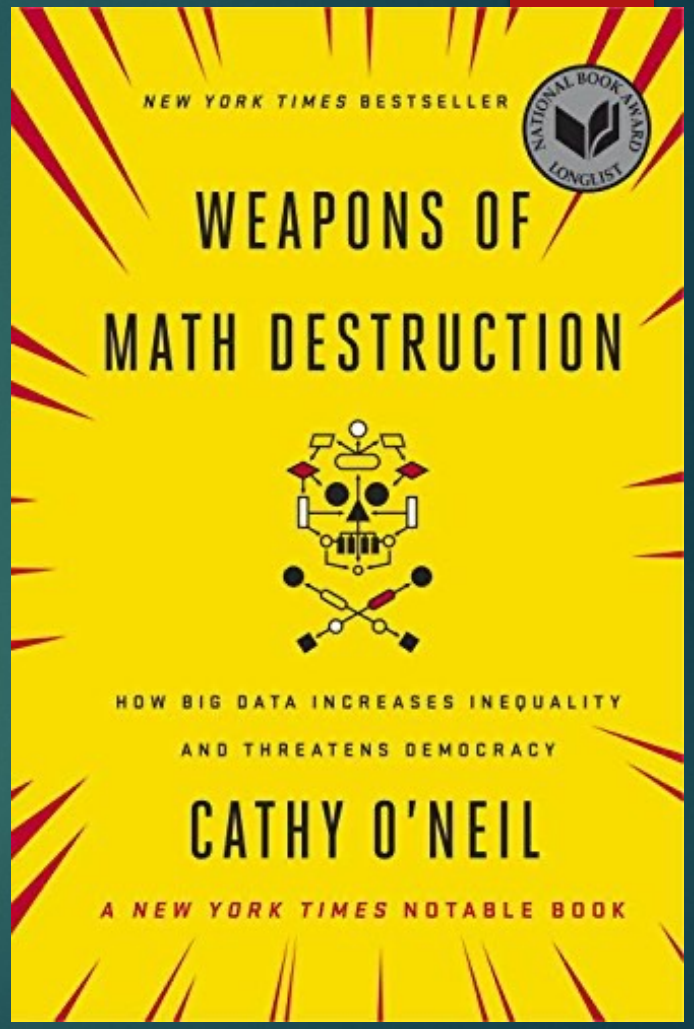
- In 2014, Amazon developed an AI system aiming at streamlining the company's hiring process by sifting through resumes and identify promising candidates for employment.
- The project was abandoned later due to the discovery that the algorithm was biased against female applicants for technical roles, like software engineering.

Bias against minority groups is not a new issue

- People blamed AI for this mistake, but this bias predates AI technology.
- In the late 1970s and early 1980s, St George Hospital Medical School in London introduced a system aimed at identifying suitable candidates for interviews.
- It was subsequently discovered that the algorithms used by this system were inclined to assign lower scores to women and racial minorities.
- Does AI tend to have more bias? Will it be worse if we rely on human ratings or pre-AI algorithms?

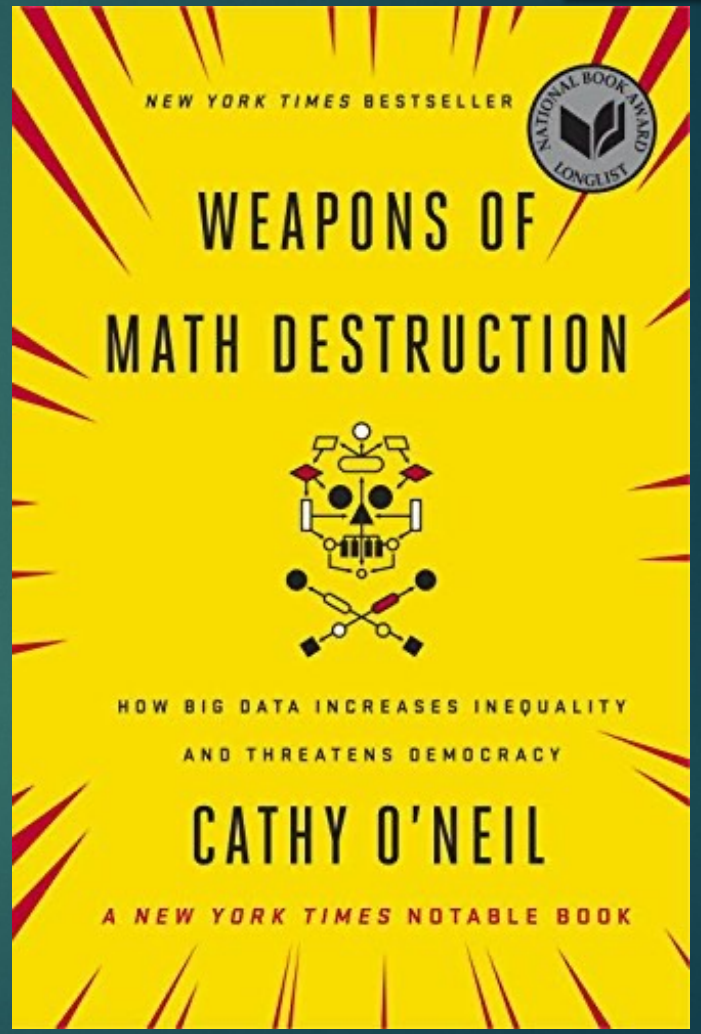
Over-focus on poor examples but overlook successful cases

- Weapon of Math Destruction has gained significant popularity. It primarily focuses on the implications of big data, a topic that is increasingly relevant as AI and big data become more intertwined.
- While the book highlights numerous instances where the misuse of data leads to negative outcomes, it often overlooks the positive applications and success stories of big data and AI.



Over-focus on poor examples but overlook successful cases

- Bad example: An individual shares the same name and the same birthday with a convict, and thus faced unfair obstacles in employment. This scenario underscores the need for enhanced data management and improved algorithms rather than a reduction in their use.
- We need more data and more math, not less!



Are the problems fixable?

- Every innovation has advantages and disadvantages.
- The crucial point is whether the advantages can outweigh the disadvantages.
- Most people use it on a daily basis to improve efficiency.
- Every technology faces difficulties when it is still in the development stage.
- The key question is: Can the problems be fixed?

Are the problems fixable?

- In 2018, Joy Buolamwini, a Black graduate student at MIT, discovered significant bias in facial recognition technology: It fails to accurately identify her dark-skinned face.
- The software easily detected the faces of people with lighter skin but failed to recognize hers until she wore a white mask.

Are the problems fixable?

- Buolamwini's work led her to found the Algorithmic Justice League (AJL), an organization committed to highlighting and mitigating the harms of AI bias.
- Her research, including the influential Gender Shades project co-authored with Timnit Gebru, has spurred tech companies to reevaluate and improve their facial recognition technologies.
- Facial recognition systems are far from perfect, but the problems can be fixed.

Are the problems fixable?

- Buolamwini observed that in Stable Diffusion, a text-to-image generative AI system, inputs prompting for high-paying professions predominantly produced images of light-skinned men.
- When the users request images of drug dealers, terrorists, or inmates, the generated images often portrayed men with darker skin.

Are the problems fixable?

- This issue may not stem directly from the algorithms themselves but rather from pre-existing biases within the datasets used to train these systems. These biases are reflective of historical and societal prejudices rather than a new phenomenon introduced by AI.
- Addressing this challenge might involve adjusting the datasets to ensure a more balanced representation, which could help mitigate the bias present in the output of AI systems.

Are the problems fixable?

- I/O psychologists Landers and Behrend (2022) argued that AI-empowered predictive models for high-stake decision support must go through rigorous and thorough audits.
- The auditing process needs to involve multiple auditors, including internal auditors, external auditors from consulting firms, and independent auditors from a regulatory authority.
- The criteria should encompass various aspects of the AI system, including data quality, model design, model development, model validation, generalizability, cultural context, and many others.

No solution is 100% fool-proof

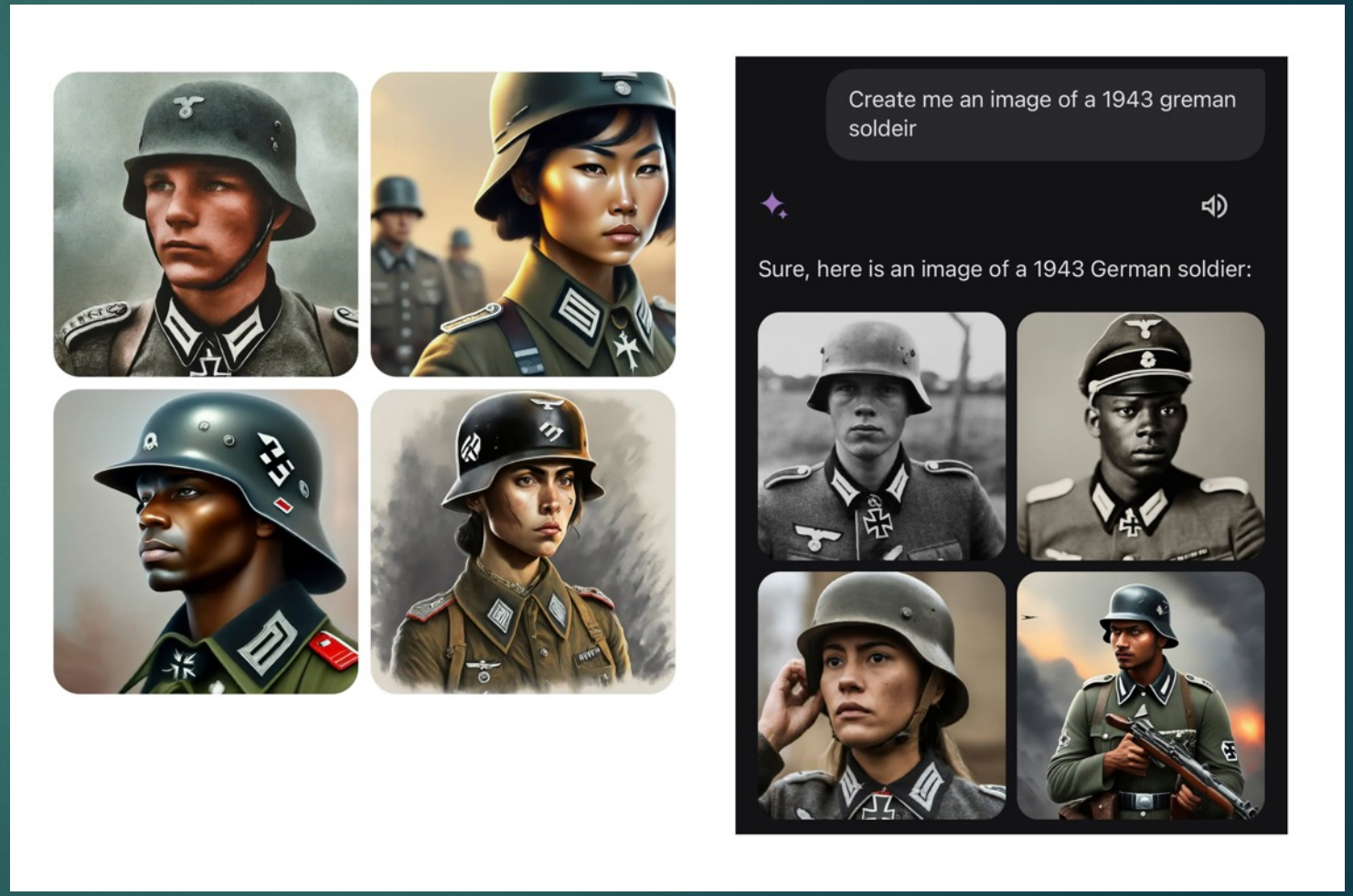
- While efforts can be made to gather more inclusive data, this initiative might conflict with the ethical principle of privacy and confidentiality.
- There has been widespread concern over agencies, organizations, and corporations collecting or even selling personal data without consent.
- Consequently, many governments have implemented regulations to curb data collection practices.

No solution is 100% fool-proof

- Artists have voiced concerns that generative art tools, such as Midjourney and Stable Diffusion, appropriate their work.
- In response, there is now an option for artists to opt out, preventing web crawlers from archiving their images.
- This method of opting in and out may result in self-selected samples that are not fully representative of the broader population.
- There is no perfect solution. We need to set a realistic goal.

Should we also worry about over-correction?

- Google Gemini, in an attempt to address AI biases concerning race and gender, produced images that were factually incorrect.
- In Feb 2024 when users entered "1943 German soldier", it yielded images that included black and Asian female soldiers.



Similar problems were seen with prompts that resulted in black Vikings, a female pope, women in the NHL, the Google founders as Asian men, and non-white depictions among the U.S. Founding Fathers.

Should we also worry about over-correction?

- Francis Bacon: "Sometime the solution is worse than the problem."

Summary

- Bias against minority groups is not a new issue.
- Do we over-focus on poor examples but overlook successful cases?
- Are the problems fixable?
- No solution is 100% fool-proof.
- Should we also worry about over-correction?

Q & A

- Contact info:
  - cayu@hpu.edu
  - chonghoyu@gmail.com