Icebreaker

Please share in the chat: What algorithms have you interacted with today?

Breaking the Code: Understanding Algorithmic Bias

Science Research Workshop Series

Slides derived & inspired by: https://libguides.lmu.edu/digcitizen/algobias ; https://sandbox.acrl.org/library-collection/foryou-algorithms-attention-economy

Welcome!

Shelby Hallman

- Physical Sciences & Engineering Librarian

Ashley Peterson

- Research & Instruction Librarian, Media & Data Literacy

Alexandra Solodkaya

- Rothman Family Food Studies Librarian

Image courtesy of Pexels

What are we doing today?

- Define algorithms
- Reflect on the role of algorithms in daily life
- Identify the potential causes of algorithmic bias and some preventive strategies

Algorithm Spectrum Activity

Assigning algorithms to a spectrum of human intervention versus automated

bit.ly/algo_spectrum

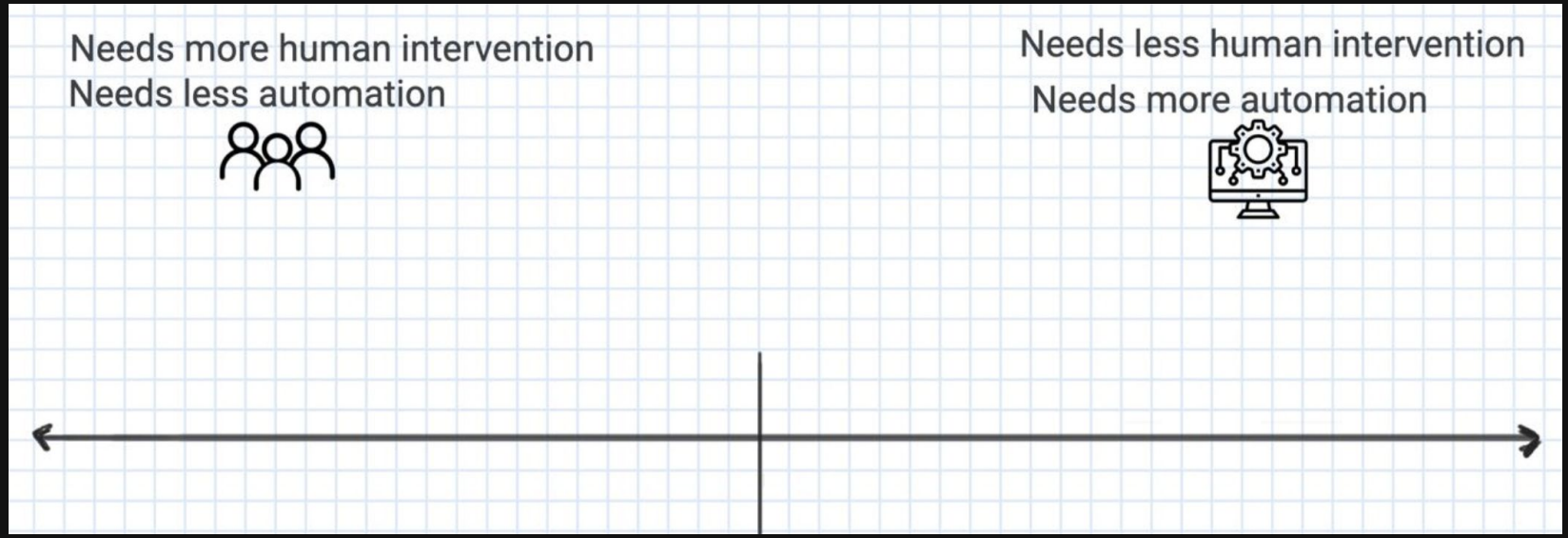
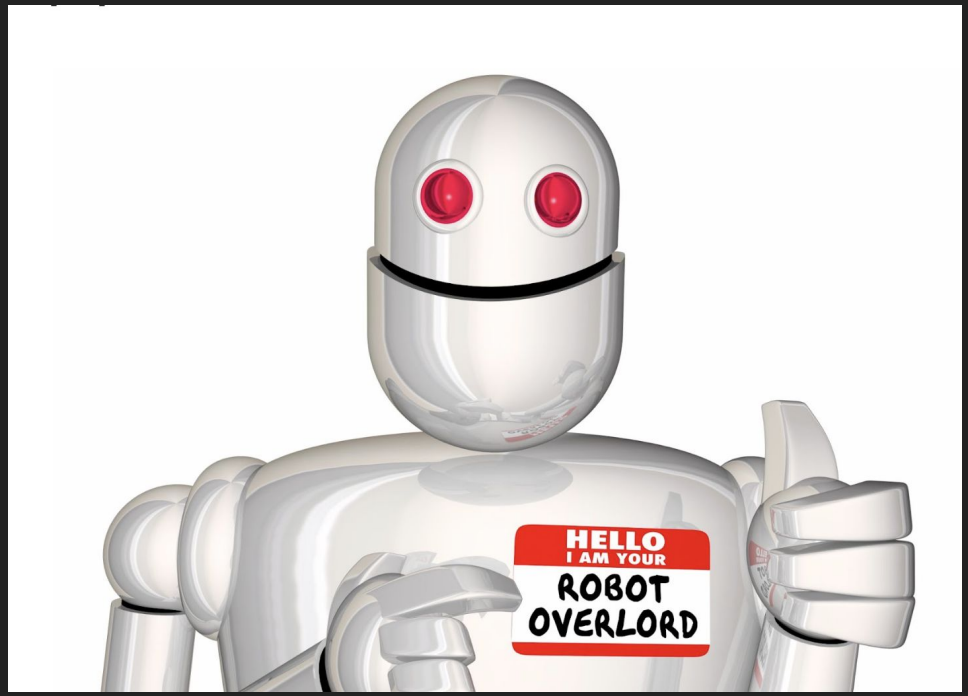


Image courtesy of Unsplash

Algorithms, AI & the end of the world (?)



What is an algorithm?

"[Algorithms]...are mathematical objects. They take a sequence of mathematical operations...and translate them into computer code. They are fed with data from the real world, given an objective and set to work crunching through the calculations to achieve their aim."

Fry, H. (2018). *Hello world: Being human in the age of algorithms*. W. W. Norton & Company

See also: (Golbeck, 2016)

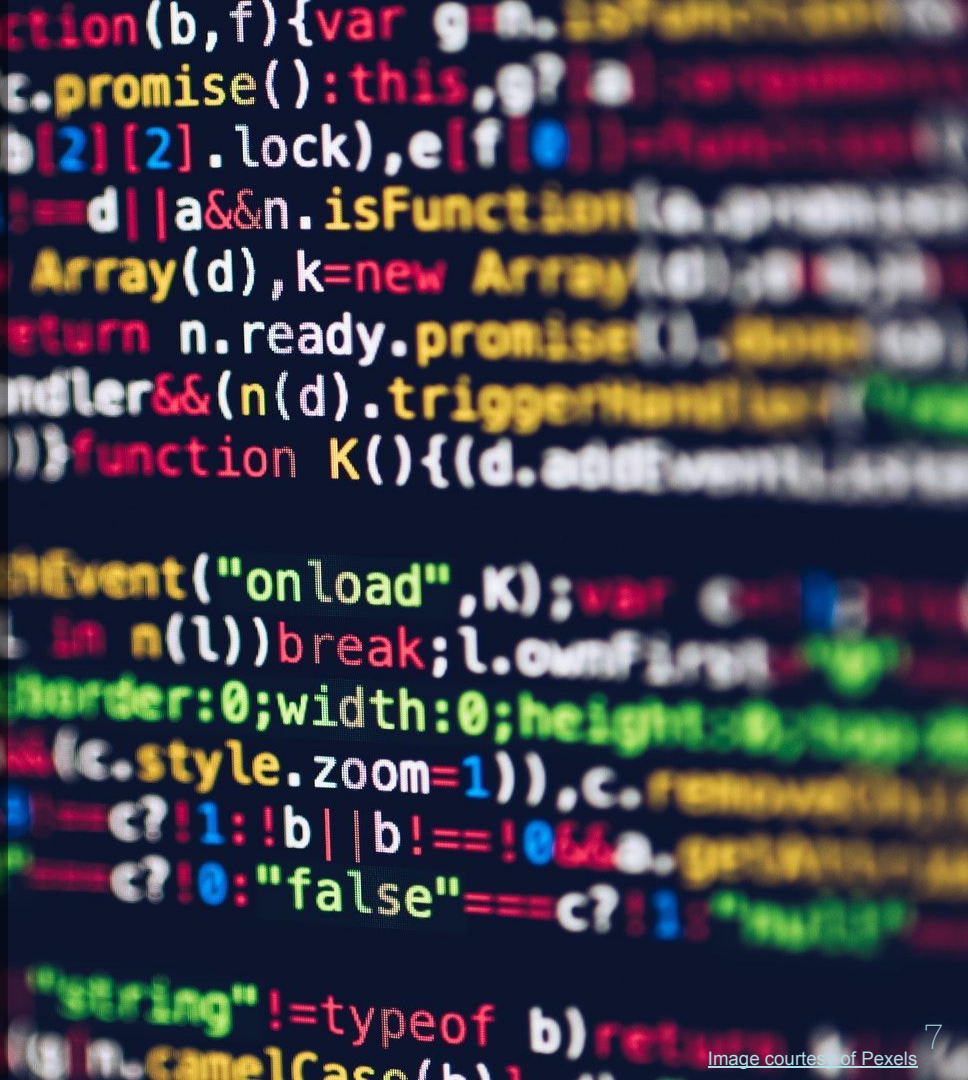


Image courtesy of Pexels

Types of algorithms

Rule-based algorithms

A set of human-coded rules that result in pre-defined outcomes (e.g., if X performs Y, then Z is the result).

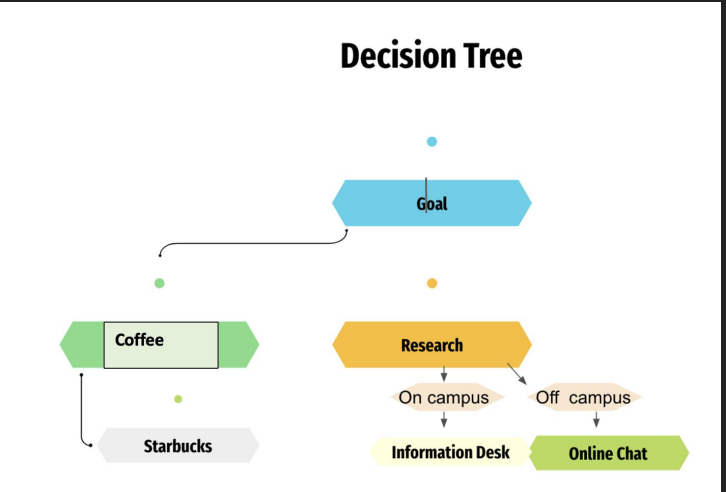
Machine-learning algorithms

Programmed to define its own set of rules, without human intervention. Output based on statistical analysis of large data sets.

Source: Grunitz, M. (2021, September 13). Rule-based AI vs machine learning: What's the difference? *WeAreBrain*. https://wearebrain.com/blog/ai-data-science/rule-based-ai-vs-machine-learning-whats-the-difference/

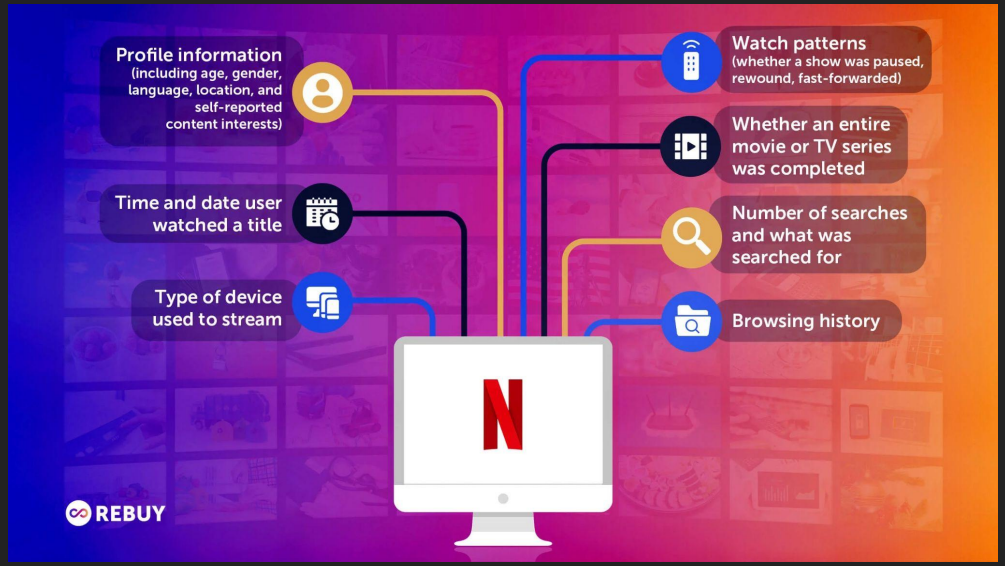
Example: Rule-based algorithm

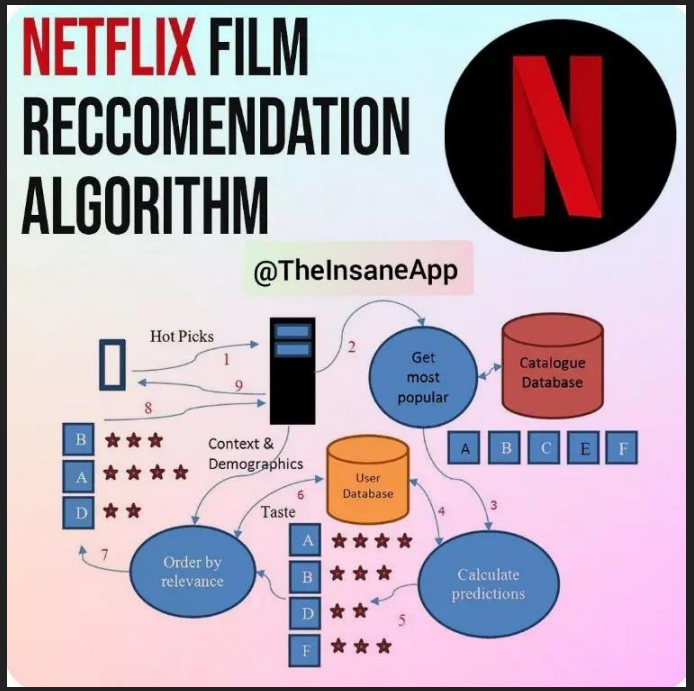
```
# in this program, we are encoding a person's options navigating the library

goals = ['coffee', 'research']
locations = ['on campus', 'off campus']

def navigate_library(goals, locations=0):
    if goals == 'coffee' and actions == 0:
        location = 'Starbucks'
            message = "Have some Starbucks coffee!"
    elif goals == 'research':
        if locations == 'on campus':
            message = "Visit us at the Information Desk!"
        elif locations == 'off campus':
            message = "Use our ask-a-librarian chat service."
```



Adapted Clark, J. A. (2018). jasonclark/algorithmic-awareness [Rich Text Format]. GitHub. https://github.com/jasonclark/algorithmic-awareness

Example: Machine learning algorithm





A note on algorithmic infrastructure

Source: https://www.invisibly.com/learn-blog/netflix-recommendation-algorithm/

Algorithmic Bias

Image courtesy of Unsplash

Algorithmic Bias

Decision-making by computer systems that delivers outcomes that are systematically less favorable to individuals within a particular group and where there is no relevant difference between groups that justified such harms. (www.unwomen.org)

Poll

What causes bias in algorithms?

a) Historical human biases in training datasets

b) Incomplete or unrepresentative training data

c) Proxies for sensitive attributes become feedback loops

d) Algorithmic objectives

Image courtesy of Unsplash

Causes of Bias

Historical Biases

Human biases included in training datasets

Unrepresentative Training Data

Incomplete or unrepresentative data

Proxies & Feedback Loops

Proxies for sensitive attributes become loops

Algorithmic Objectives

Minimize prediction errors and benefit majority groups

Example- Historical Bias

Microsoft's Tay Twitter Chatbot

- Trained on anonymized public data from Internet
- After 1 day, Tay had to be shut down for a series of lewd and racist tweets

Image credit: Vincent, J. (2016, March 24) Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge.* https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Data Labelers Remove Toxic Content in ChatGPT

Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

Image Credit: Perrigo, B. (2023, January 18). OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic. *Time*. *https://time.com/6247678/openai-chatgpt-kenya-workers/*

Example–Unrepresentative Training Data

Facial Recognition Software

- Training set mostly lighter-skinned faces
- Darker-skinned females were the most misclassified group (error rates of up to 34.7%)
- Maximum error rate for lighter-skinned males at 0.8%.

Buolamwini, J., & Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91.

Image credit: Buolamwini, J. (2016, Dec. 14). T*he Algorithmic Justice League*. MIT Media Lab. https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148.

Example–Proxies and Feedback Loops

COMPAS Recidivism Algorithm

- Brisha, a black teenager with a previous misdemeanor who stole a kid's bike, was assigned a "high" risk
- Vernon, a white man who shoplifted goods equivalent to the value of the bike but had a previous criminal record, was a "low" risk

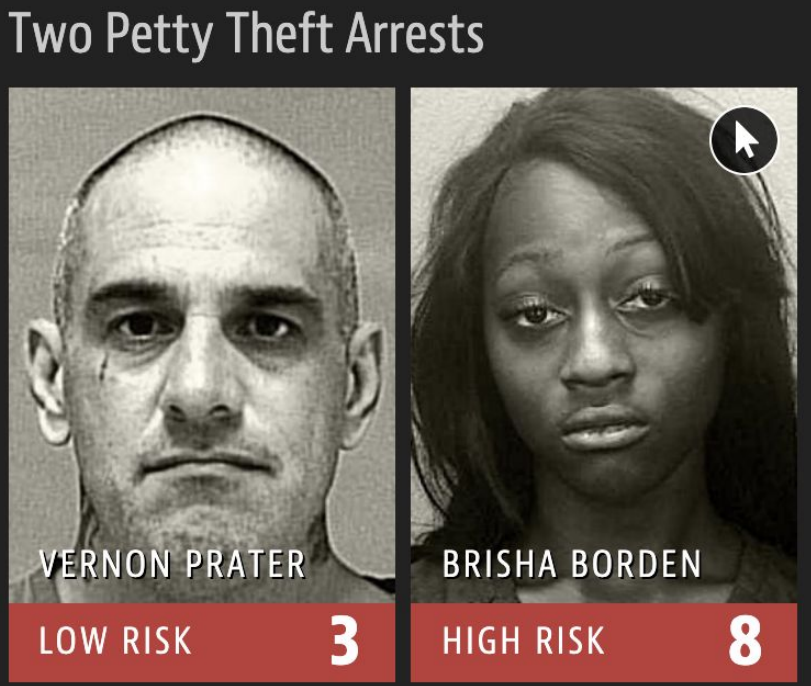


Image Credit: Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=aGn0cAzF1vcRrdeL0ElVR2hgvgtEPpJo

Example–Algorithmic Objectives

Healthcare Algorithm

- Used to diagnose patients with bipolar disorder
- Uses strict DSM-5 criteria
- Prioritizes accuracy over context
- Would not intentionally misdiagnose, even if better for the patient

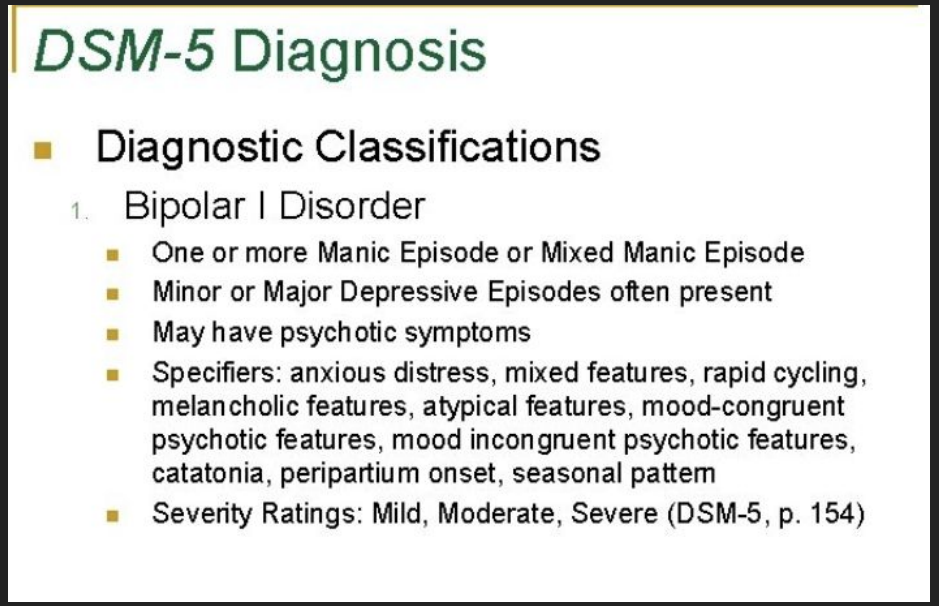


Image credit: American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association Pub.

Sneak peek: Bias in generative AI tools

How to build a Large Language Model (LLM)

1. Set a goal
2. Collect data (lots!), tokenize it
3. Build a neural network
4. Train the neural network
5. Fine-tune the model
6. Build a user interface & launch

Adapted from How to Become an Expert on A.I., by Kevin Roose and Cade Metz for the New York Times

How might bias manifest in any of these steps?

Algorithm Spectrum Activity

Assigning algorithms to a spectrum of human intervention versus automated

bit.ly/algo_spectrum

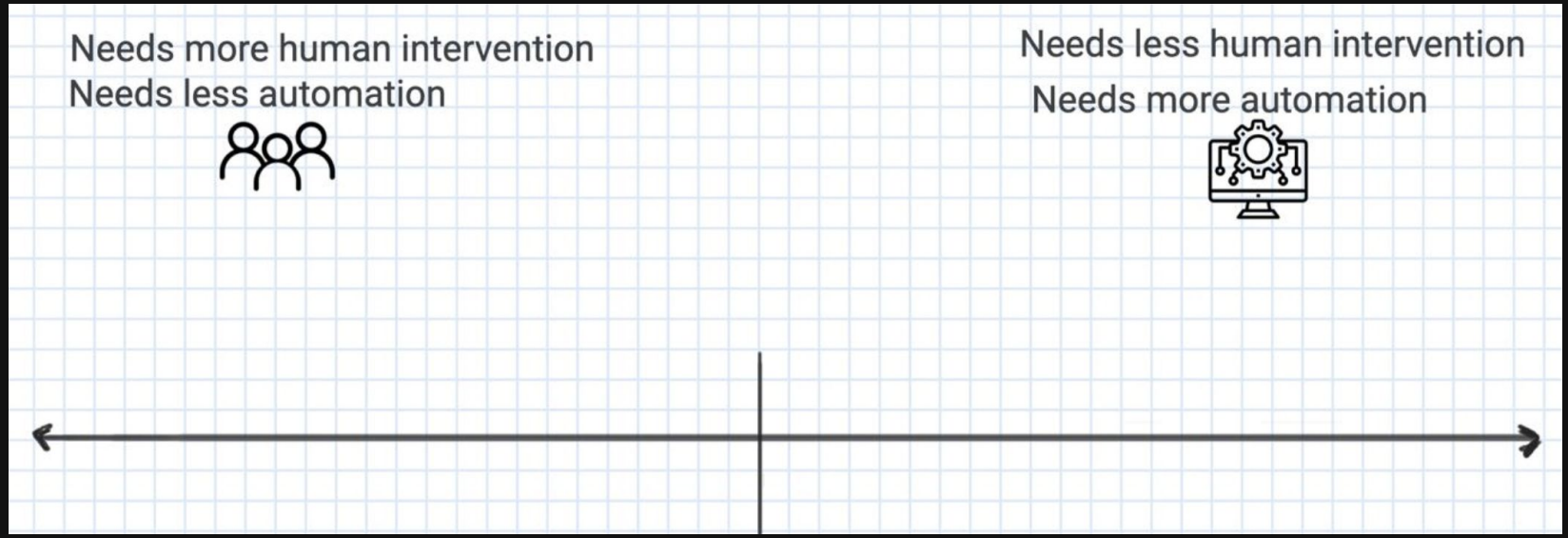


Image courtesy of Unsplash

Our Takeaways

Think Twice 

Before downloading free apps - you are paying with your data!

Provide Feedback 

Biased or inappropriate search results? Provide search feedback

Request Transparency 

From corporations

Opt-out 

From data collection whenever possible

Get Involved 

Advocacy and educational groups

Get Involved!

- Our Data Bodies
  - https://www.odbproject.org/
- Electronic Frontier Foundation
  - https://www.eff.org/
- Carceral Tech Resistance Network
  - https://www.carceral.tech/
- Detroit Community Technology Project
  - https://detroitcommunitytech.org/
- Data & Society
  - https://datasociety.net/
- AI Now Institute
  - https://ainowinstitute.org/

Image courtesy of Pexels

Further Reading

- The Supremacy of Bias in AI
- The AI takeover of Google Starts Now
- The Problem With Biased AIs (and How To Make AI Better)
- "Do We Need Librarians Now that We Have ChatGPT?"
- AI and the future of information literacy and information ethics
- The dilemma of the direct answer
- How to Become an Expert on A.I.
- Chatbots could one day replace search engines. Here's why that's a terrible idea.
- OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic
- AI Fairness 360
- Noble, Safiya Umoja. 2018. Algorithms of Oppression : How Search Engines Reinforce Racism
- O'Neil, Cathy. 2016. Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy
- Broussard, Meredith. 2018. Artificial Unintelligence : How Computers Misunderstand the World
- Srinivasan, Ramesh. 2017. Whose Global Village? : Rethinking How Technology Shapes Our World
- Nemer, David. 2022. Technology of the Oppressed: Inequity and the Digital Mundane in Favelas of Brazil
- Roberts, Sarah T. 2019. Behind the Screen : Content Moderation in the Shadows of Social Media

Evaluation

Thanks!

Science Research Workshop Series
UCLA Library

Slides derived & inspired by: https://libguides.lmu.edu/digcitizen/algobias ; https://sandbox.acrl.org/library-collection/foryou-algorithms-attention-economy

Image courtesy of Unsplash