8-26-2024

Psychology and the Digital Everywhere: Artificial Intelligence

Cassie van Stolk Assistant Professor of Psychology, Department of Psychology
SUNY Geneseo, cvanstolk@geneseo.edu

Recommended Citation

IDEAS THAT MATTER

2024-2025 Artificial Intelligence

AI – A New Frontier
Class Activity: Case Study

Prepared by Cassie van Stolk, Assistant Professor of Psychology, Department of Psychology

Case Study: *Is This AI System Biased?*

Instructions:
Students will analyze a case study of an AI system, considering how it was developed and how it continues to "learn." In small groups, students will discuss the degree to which they feel the AI system might be biased, discuss implications of bias, and propose strategies to mitigate bias and its psychological consequences to AI users. Students will collaborate to complete a worksheet addressing these components.

Grading Rubric:

Criteria	Unsatisfactory Excellent				
	1	2	3	4	5
Required Components					
Recognizing Bias: Teams will identify examples of bias that arise in the case study, and provide explanations for how they believe bias was originated and perpetuated.	1	2	3	4	5
Critical Analysis: Teams will analyze the potential impacts of bias on affected individuals or groups using an ecological framework that considers intrapersonal, interpersonal, and societal consequences.	1	2	3	4	5
Solution Proposal: Teams will propose practical strategies to mitigate the consequences of the bias present within the AI system. Justification of solutions will be based on psychological and ethical principles covered in class.	1	2	3	4	5
Team Collaboration: Teams will effectively collaborate and communicate with each other. All members will equally contribute. Presentation of responses on the team worksheet should be clear and coherent.	1	2	3	4	5
Effective Communication: Presentation of responses on the team worksheet should be clear and coherent, representing the contributions of all members.	1	2	3	4	5

GENESEO
THE STATE UNIVERSITY OF NEW YORK

Case Study: Team Worksheet

Read your assigned case study aloud. Team members should break the case study into sections, with a reader assigned to each section. Members who are not reading a section should be listening actively.

Once you have read your case study, answer the following questions:

1. Did your team notice any examples of bias within the AI system that you read about? What types of bias did you notice? Please use psychological terminology for bias covered in class.

   _____

   _____

   _____

   _____

   _____

2. How do you think that bias **originated** within the AI system? How is it being **perpetuated**?

   _____

   _____

   _____

   _____

_____

_____

3. Using an ecological framework, identify potential consequences of AI bias to the individual, immediate interpersonal communities, and society more broadly:

Individual (consider physiological, cognitive, and behavioral consequences)

_____

_____

_____

_____

_____

Interpersonal (consider immediate relationships among friends, family, neighbors, communities)

_____

_____

_____

_____

Societal (consider media, industry, politics, faith, and culture)

_____

_____

_____

_____

_____

4. Identify 1-2 possible solutions that are designed to mitigate (lessen) the negative consequences of bias derived from principles of ethics and psychological well being.

_____

_____

_____

_____

_____

Facial Recognition Software in Criminal Justice

In bustling Metropolis, a team of engineers and data scientists embarked on a groundbreaking mission to revolutionize law enforcement. Led by a renowned expert in artificial intelligence, the team set out to develop cutting-edge facial recognition software that could swiftly identify suspects of urban crimes, with the goal of solving crimes more quickly and accurately.

Months of research and development culminated in the creation of *FaceGuard*, an advanced AI system trained that used state-of-the-art algorithms and deep learning techniques to train on vast datasets of facial images collected from a wide range of sources, including public databases, social media platforms, government records, and law enforcement archives. While an effort was made to collect facial images from diverse ethnic and cultural groups, it was not always possible to get sufficient data representing certain types of racial and ethnic communities.

As *FaceGuard* deployed across Metropolis' law enforcement agencies, it quickly garnered attention for its remarkable speed and efficiency in identifying suspects from surveillance footage and police databases. Detectives hailed it as a game-changer, praising its ability to match faces with unparalleled accuracy, leading to swift arrests and convictions.

The team who had developed *FaceGuard* monitored its use, with a particular focus on assessing its face-matching accuracy. The more it was used, the more the team noticed a patter: *FaceGuard's* accuracy statistics were consistently emerging as significantly lower for individuals from certain racial or ethnic groups, particularly those with darker skin tones.

Examining the data on how *FaceGuard* was being used by law enforcement, the team further noticed that the majority of arrests and convictions being made using *FaceGuard* identified individuals of color – and particularly individuals with darker skin tones – as the perpetrators. Furthermore, as AI technology, FaceGuard was still learning, primarily using each agency's newest arrest and conviction data to train on faces of identified perpetrators.

Outlines for Additional Case Studies

Automated Resume Screening Tool
- Description: A company implements an automated resume screening tool to filter job applicants. However, the tool consistently rejects resumes from applicants with names commonly associated with minority groups.

- Bias: The algorithm is trained on historical hiring data, which reflects existing biases and discrimination in the workforce.

- Consequences: Qualified candidates from underrepresented backgrounds are unfairly excluded from job opportunities, perpetuating systemic inequalities in employment.

Healthcare Algorithm for Predicting Patient Risk
- Description: A healthcare algorithm is developed to predict patient risk levels for various medical conditions. However, the algorithm systematically underestimates the risk for patients from low-income neighborhoods.

- Bias: The algorithm relies heavily on insurance claims data, which may not accurately capture the healthcare needs and access barriers faced by individuals from marginalized communities.

- Consequences: Patients from disadvantaged backgrounds are less likely to receive timely interventions or preventive care, exacerbating health disparities.

Automated Loan Approval System
- Description: A bank introduces an automated system for processing loan applications, aiming to streamline the approval process. However, the system consistently approves loans for applicants from affluent neighborhoods while denying loans for applicants from economically disadvantaged areas.

- Bias: The algorithm incorporates zip codes or neighborhood data as predictors of creditworthiness, reinforcing socioeconomic biases.

- Consequences: Qualified borrowers from low-income communities are unfairly denied access to financial resources, hindering their ability to secure housing, education, or entrepreneurship opportunities.