



Welcome to this session

Skills Bootcamp:

Case Studies into Data Science Ethics

The session will start shortly...

Questions? Drop them in the chat.
We'll have dedicated moderators
answering questions.



Skills Bootcamp Data Science Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(Fundamental British Values: Mutual Respect and Tolerance)**
- No question is daft or silly - **ask them!**
- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. We will be answering questions as the session progresses as well.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Academic Sessions. You can submit these questions here: **Questions**

Skills Bootcamp Data Science Housekeeping

- For all **non-academic questions**, please submit a query: www.hyperiondev.com/support
- Report a safeguarding incident: www.hyperiondev.com/safeguardreporting
- We would love your feedback on lectures: [Feedback on Lectures.](#)
- Find all the lecture **content** in your [Lecture Backpack](#) on GitHub.
- If you are hearing impaired, kindly use your computer's function through Google chrome to enable captions.

Safeguarding & Welfare

We are committed to all our students and staff feeling safe and happy; we want to make sure there is always someone you can turn to if you are worried about anything.

If you are feeling upset or unsafe, are worried about a friend, student or family member, or you feel like something isn't right, speak to our safeguarding team:



Ian Wyles
Designated Safeguarding
Lead



Simone Botes



Nurhaan Snyman



Rafiq Manan



Ronald Munodawafa



Tevin Pitts

Scan to report a
safeguarding concern



or email the Designated
Safeguarding Lead:
Ian Wyles

safeguarding@hyperiondev.com

Skills Bootcamp Progression Overview

✓ Criterion 1 - Initial Requirements

Specific achievements **within the first two weeks** of the program.

To meet this criterion, students need to, by no later than **01 December 2024 (C11)** or **22 December 2024 (C12)**:

- **Guided Learning Hours (GLH):** Attend a **minimum of 7-8 GLH per week** (lectures, workshops, or mentor calls) for a total minimum of **15 GLH**.
- **Task Completion:** Successfully complete the **first 4 of the assigned tasks**.

✓ Criterion 2 - Mid-Course Progress

Progress through the successful completion of tasks **within the first half** of the program.

To meet this criterion, students should, by no later than **12 January 2025 (C11)** or **02 February 2025 (C12)**:

- **Guided Learning Hours (GLH):** Complete at least **60 GLH**.
- **Task Completion :** Successfully complete the **first 13 of the assigned tasks**.

Skills Bootcamp Progression Overview

✓ Criterion 3 – End-Course Progress

Showcasing students' progress nearing the completion of the course.

To meet this criterion, students should:

- **Guided Learning Hours (GLH):** Complete the **total minimum required GLH**, by the **support end date**.
- **Task Completion : Complete all mandatory tasks**, including any necessary resubmissions, by the end of the bootcamp, **09 March 2025 (C11)** or **30 March 2025 (C12)**.

✓ Criterion 4 - Employability

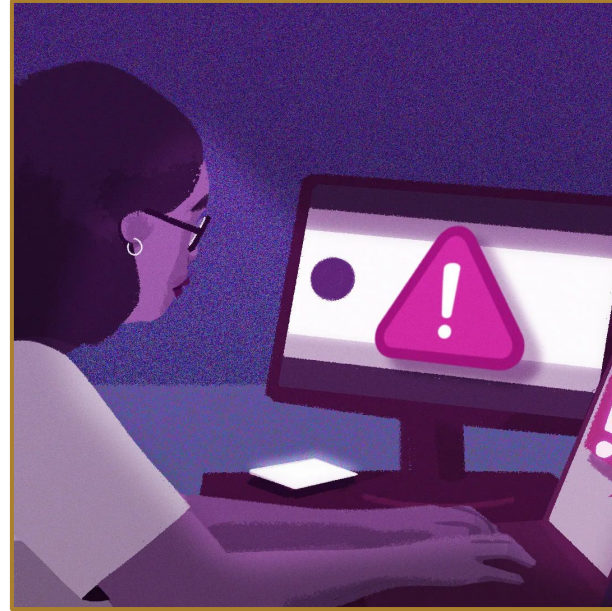
Demonstrating progress to find employment.

To meet this criterion, students should:

- **Record an Interview Invite:** Students are required to record proof of invitation to an interview by **30 March 2025 (C11)** or **04 May 2025 (C12)**.
 - **South Holland Students** are required to proof and interview by **17 March 2025**.
- **Record a Final Job Outcome :** Within 12 weeks post-graduation, students are required to record a job outcome.

Question

- ❏ Can you think of a time when a technology made a decision that felt unfair or biased to you or someone you know? What happened?



Learning Outcomes

- ❖ Identify Ethical Issues in AI Systems
- ❖ Analyze Case Studies of Ethical AI Failures and Successes
- ❖ Apply Ethical Frameworks to AI Development
- ❖ Propose Solutions to Ethical Challenges

Lecture Overview

- Historical cases of ethical failures
- Success stories
- Lessons learned and best practices
- Discussion





Introduction to Case Studies in Data Science Ethics

Introduction

- ❖ Data science has transformed industries, but ethical challenges often arise, impacting individuals and communities. This lecture explores historical failures, success stories, and key lessons in ethical AI development.

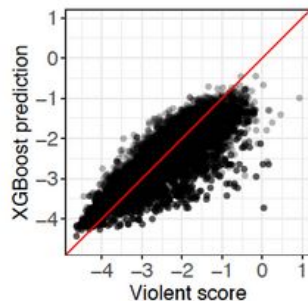


Historical Cases of Ethical Failures

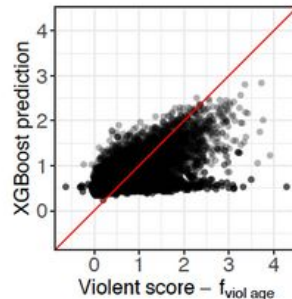


1.1. COMPAS: Bias in Criminal Justice

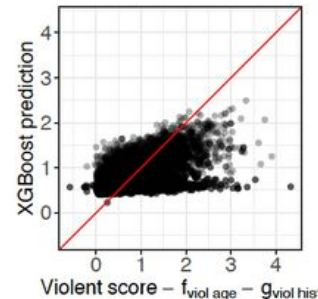
- ❖ **Issue:** The COMPAS algorithm, used in the U.S. to assess criminal recidivism risk, disproportionately classified Black defendants as high risk.



(a) Predictions of COMPAS violent score vs. actual values.



(b) Predictions of COMPAS violent score $-f_{\text{viol age}}$ vs. actual values.



(c) Predictions of COMPAS violent score $-f_{\text{viol age}} - g_{\text{viol hist}}$ vs. actual values.

Ethical Failures

- ❖ Bias in training data leading to unfair predictions.
- ❖ Lack of transparency in decision-making.
- ❖ No accountability for algorithmic errors.
- ❖ **Impact:**
 - Led to wrongful sentencing disparities and loss of public trust.



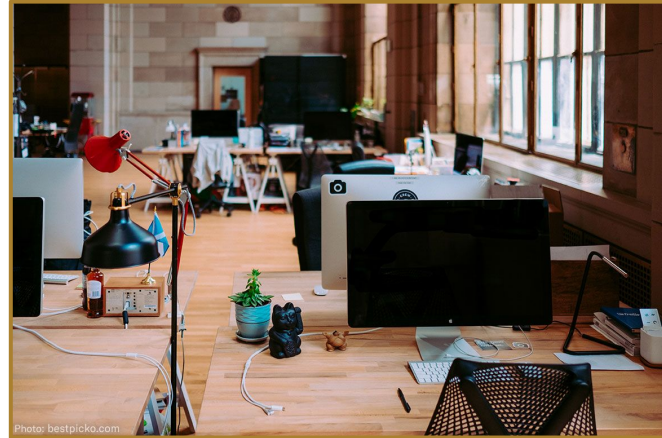
1.2. Amazon's Discriminatory Hiring Algorithm

- ❖ **Issue:** Amazon trained a hiring algorithm that downgraded female candidates.



Ethical Failures

- ❖ Training data was based on past (biased) hiring patterns.
- ❖ Lack of checks for bias before deployment.
- ❖ The model perpetuated gender discrimination.
- ❖ **Impact:**
 - Reinforced systemic biases in hiring processes.



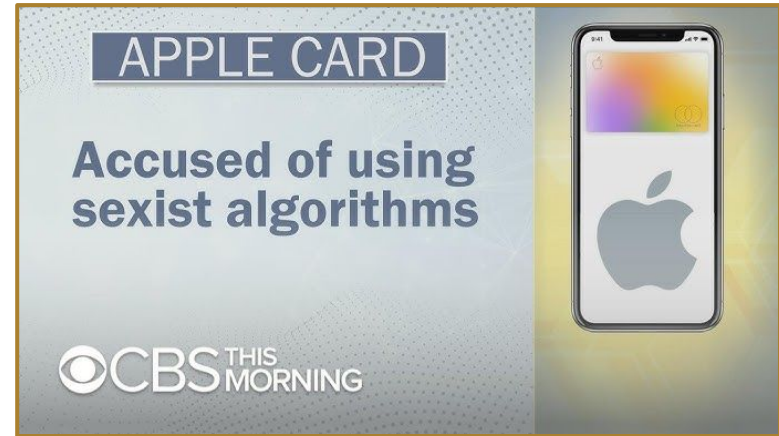
1.3. Apple Card: Gender Bias in Credit Decisions

- ❖ **Issue:** Apple's credit scoring algorithm gave significantly lower credit limits to women compared to men with identical profiles.



Ethical Failures

- ❖ Lack of transparency in credit decision-making.
- ❖ No clear appeal process for affected users.
- ❖ Failure to conduct fairness testing before release.
- ❖ **Impact:**
 - Reinforced financial discrimination.



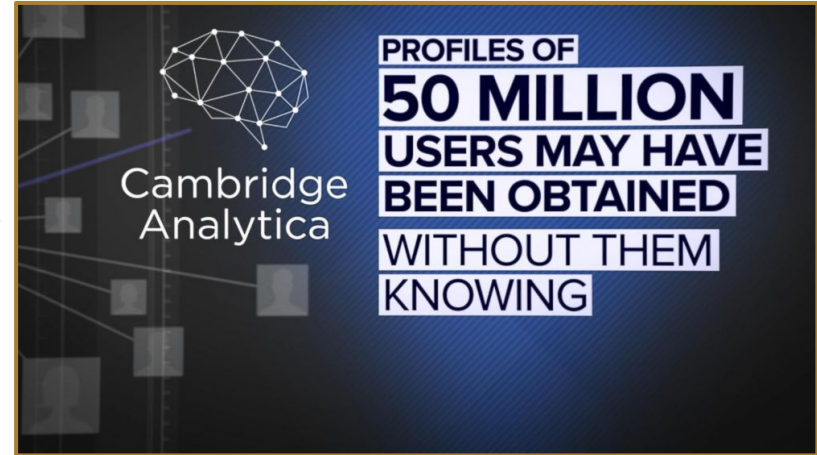
1.4. Cambridge Analytica & Data Privacy Violations

- ❖ **Issue:** Cambridge Analytica harvested Facebook user data without consent to influence elections.



Ethical Failures

- ❖ Unauthorized data collection.
- ❖ Manipulative political targeting.
- ❖ Violation of user privacy.
- ❖ **Impact:**
 - Global political consequences and decline in trust in social media platforms.

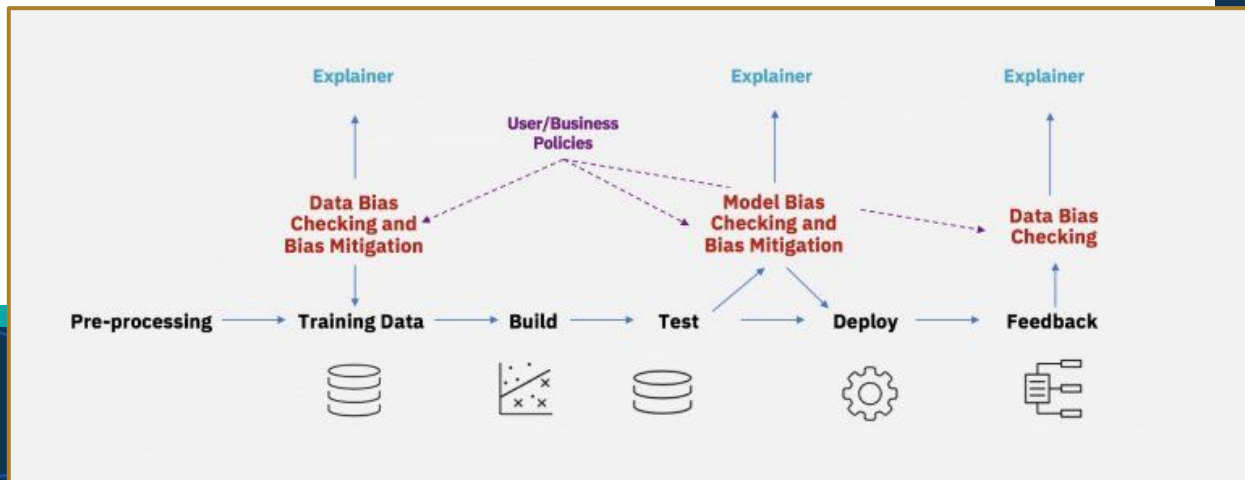




2. Success Stories: Ethical AI Initiatives

2.1. IBM's AI Fairness 360 Toolkit

- ❖ What it does:
 - Provides open-source tools to detect and mitigate AI bias.
- ❖ Success Factors:
 - Promotes fairness assessments.
 - Enables developers to test for bias before deployment.
 - Encourages transparency in AI decision-making.



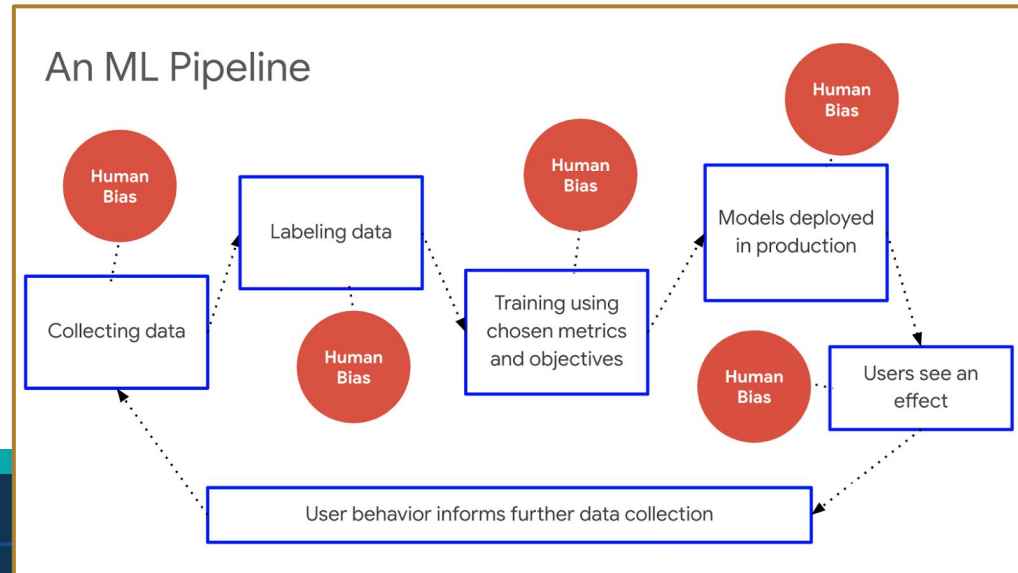
2.2. Microsoft's AI for Accessibility

- ❖ What it does:
 - Develops AI solutions for individuals with disabilities.
- ❖ Success Factors:
 - Inclusive design principles.
 - AI models tailored for accessibility needs.
 - Community-driven feedback for continuous improvements.



2.3. The Fairness Indicators by Google

- ❖ What it does:
 - Open-source tool to help machine learning teams assess fairness metrics.
- ❖ Success Factors:
 - Enables bias detection in large-scale AI systems.
 - Supports fairness-aware development practices.



2.4. GDPR and Ethical Data Practices

- ❖ What it does:
 - The General Data Protection Regulation (GDPR) enforces strict data privacy laws in the EU.
- ❖ Success Factors:
 - Strengthens user consent mechanisms.
 - Ensures accountability in data collection and usage.
 - Fosters ethical data management practices worldwide.



Let's take a
break

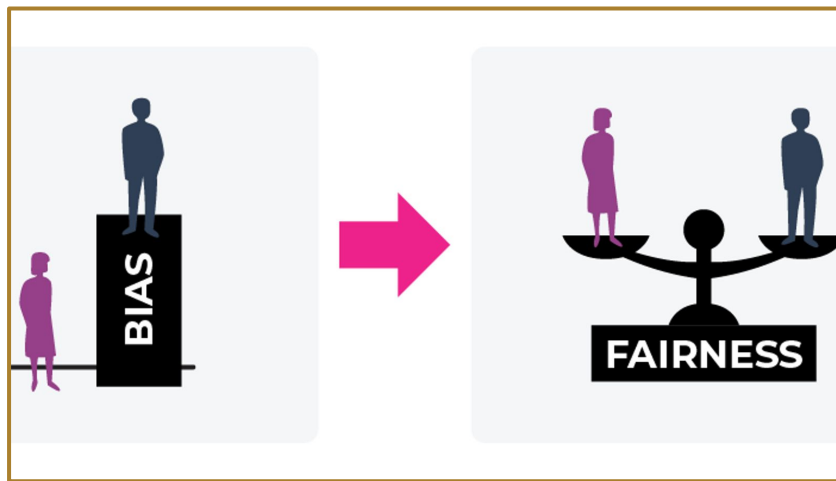




3. Lessons Learned & Best Practices

3.1. Fairness & Bias Mitigation

- ❖ Conduct fairness audits in AI models.
- ❖ Train algorithms with diverse datasets.
- ❖ Implement bias detection tools like AI Fairness 360.



3.2. Transparency & Explainability

- ❖ Use interpretable models (e.g., SHAP, LIME).
- ❖ Provide clear explanations for AI-driven decisions.
- ❖ Establish accountability frameworks.



3.3. Ethical Data Practices

- ❖ Adopt responsible data collection methods.
- ❖ Ensure informed user consent for data usage.
- ❖ Implement strict data security measures.



3.4. Regulation & Governance

- ❖ Comply with GDPR, CCPA, and other data regulations.
- ❖ Establish ethical review boards for AI projects.
- ❖ Foster cross-disciplinary collaboration (ethicists, engineers, legal experts).





**Take on roles (Data Scientist,
Regulator, Affected User, etc.)**

Case Study: Social Media Algorithm Promotes Misinformation

- ❖ What ethical considerations are neglected here?
- ❖ What measures can be taken to prevent this issue?

Polls Assessment



What was the primary ethical issue with the COMPAS algorithm in criminal justice?

- A. It lacked transparency in how risk scores were assigned.
- B. It was too lenient on repeat offenders.
- C. It prioritized economic background over criminal history.
- D. It was open-source and publicly accessible.





What was the primary ethical issue with the COMPAS algorithm in criminal justice?

- A. It lacked transparency in how risk scores were assigned.**
- B. It was too lenient on repeat offenders.
- C. It prioritized economic background over criminal history.
- D. It was open-source and publicly accessible.



Why did Amazon's AI hiring algorithm discriminate against female candidates?

- A. The algorithm was explicitly programmed to prefer male candidates.
- B. It used biased historical hiring data as training input.
- C. Women were not applying for technical roles in sufficient numbers.
- D. It favored candidates with more social media presence.



Why did Amazon's AI hiring algorithm discriminate against female candidates?

- A. The algorithm was explicitly programmed to prefer male candidates.
- B. It used biased historical hiring data as training input.**
- C. Women were not applying for technical roles in sufficient numbers.
- D. It favored candidates with more social media presence.



How did Cambridge Analytica violate data ethics?

- A. It manipulated training data to achieve biased results.
- B. It used machine learning to track users without their knowledge.
- C. It harvested personal data without user consent and used it for political targeting.
- D. It shared data only with academic researchers.





How did Cambridge Analytica violate data ethics?

- A. It manipulated training data to achieve biased results.
- B. It used machine learning to track users without their knowledge.
- C. It harvested personal data without user consent and used it for political targeting.**
- D. It shared data only with academic researchers.



Which of the following is a key lesson from past AI failures?

- A. Transparency and explainability are crucial for ethical AI.
- B. AI models should always be complex for better accuracy.
- C. Ethical considerations slow down AI development and should be minimized.
- D. Data privacy laws are unnecessary when models are well-trained.





Which of the following is a key lesson from past AI failures?

- A. Transparency and explainability are crucial for ethical AI.**
- B. AI models should always be complex for better accuracy.
- C. Ethical considerations slow down AI development and should be minimized.
- D. Data privacy laws are unnecessary when models are well-trained.



What is a major function of IBM's AI Fairness 360 Toolkit?

- A. It helps improve algorithm speed and efficiency.
- B. It provides tools to detect and mitigate bias in AI models.
- C. It automates the decision-making process in legal cases.
- D. It replaces human oversight in AI deployments.





What is a major function of IBM's AI Fairness 360 Toolkit?

- A. It helps improve algorithm speed and efficiency.
- B. It provides tools to detect and mitigate bias in AI models.**
- C. It automates the decision-making process in legal cases.
- D. It replaces human oversight in AI deployments.





Which regulation enforces strict user consent and data protection practices in the EU?

- A. CCPA
- B. GDPR
- C. HIPAA
- D. Fair Credit Reporting Act



Which regulation enforces strict user consent and data protection practices in the EU?

- A. CCPA
- B. GDPR**
- C. HIPAA
- D. Fair Credit Reporting Act



What is one effective way to mitigate bias in AI models?

- A. Use a single source of training data for consistency.
- B. Regularly audit AI models for fairness and bias.
- C. Avoid human oversight to ensure unbiased results.
- D. Increase the opacity of AI systems to protect proprietary methods.



What is one effective way to mitigate bias in AI models?

- A. Use a single source of training data for consistency.
- B. Regularly audit AI models for fairness and bias.**
- C. Avoid human oversight to ensure unbiased results.
- D. Increase the opacity of AI systems to protect proprietary methods.

Questions and Answers



Thank you for attending



CoGrammar



Department
for Education