



Welcome to this session
Skills Bootcamp:

**Ethics and Fairness in AI and Data
Science**

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.

Initial Polls Assessment

Which of the following is NOT a core ethical principle in AI?

- A. Transparency
- B. Accountability
- C. Profit Maximization
- D. Fairness



What is an example of bias in AI?

- A. A model favouring one demographic group over another
- B. A model providing consistent and fair results for all users
- C. An algorithm that adapts to user preferences
- D. A model that ignores sensitive attributes



Which regulation focuses on data protection and privacy in the European Union?

- A. CCPA
- B. GDPR
- C. HIPAA
- D. AI Act

Learning Outcomes

- Explain key ethical principles in AI and Data Science.
- Identify types of bias in datasets and machine learning algorithms.
- Describe techniques to ensure fairness in AI models.
- Discuss privacy concerns and relevant regulations.
- Discuss the importance of explainable AI (XAI) and accountability.
- Analyse real-world implications of ethical breaches in AI.

Lecture Overview

- Introduction to AI and Data Science Ethics
- Bias and Discrimination in AI
- Fairness in Algorithms
- Privacy Concerns in AI
- Accountability in AI
- Real-World Implications of Unethical AI

Introduction to Ethics and Fairness in AI and Data Science

Question

- ❖ As AI systems increasingly influence critical aspects of our lives, such as hiring, lending, and law enforcement, who should be held responsible when these systems make biased or unfair decisions? Should it be the developers, the companies deploying them, or society as a whole? How can we ensure accountability while fostering innovation?

Real-World Application of Ethics and Fairness in AI and Data Science

Imagine you're a data scientist working for a financial institution that uses AI to assess loan applications.

You have access to user data, including financial history, employment status, credit scores, and demographics. The management wants to know:

- How can we ensure our AI-driven loan approval system makes fair and unbiased decisions?
- Are there hidden biases in our dataset that could lead to discrimination against certain groups?
- How do we make our AI models transparent so that applicants understand why they were approved or denied?

Using ethical AI practices, you can audit your datasets for bias, apply fairness-aware algorithms, and implement explainable AI (XAI) techniques to ensure responsible decision-making. By prioritizing ethics, you not only comply with regulations but also build trust with users and create a more equitable financial system.

Introduction to AI and Data Science Ethics

- ❖ AI and data science have transformed industries, but with great power comes great responsibility.
- ❖ Ethics in AI ensures that these technologies are developed and deployed in ways that align with societal values.



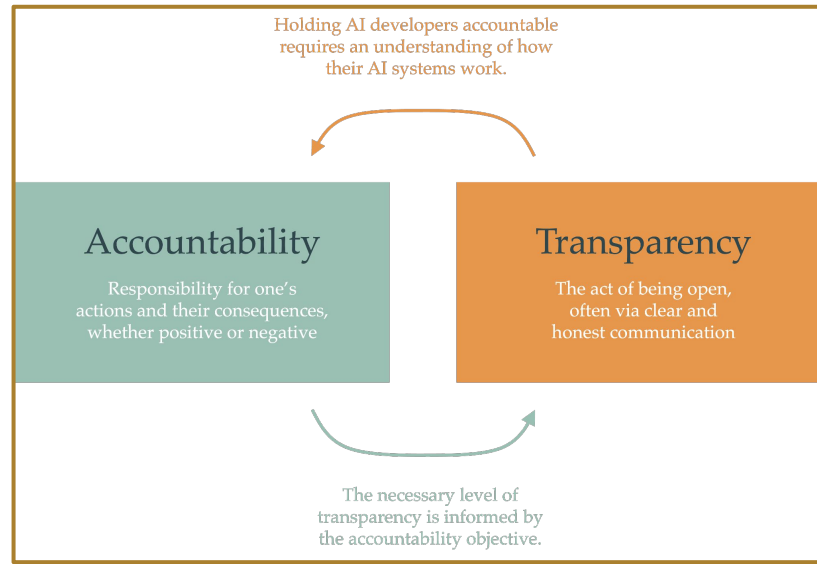
Ethical AI is built upon key principles

The 9 Pillars of **Ethical AI**



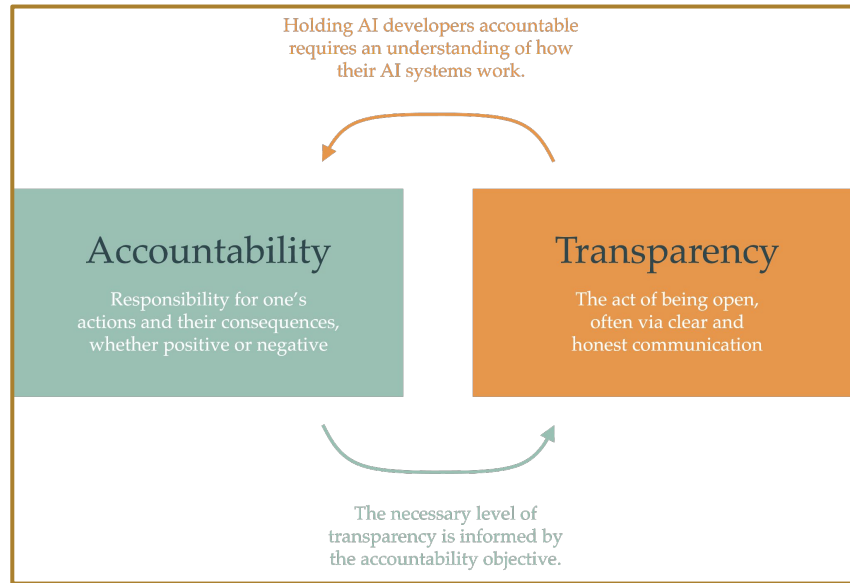
Principles of Ethical AI

Accountability: Who is responsible for AI decisions? If an AI system causes harm, can we trace back the issue and hold the right entity accountable?



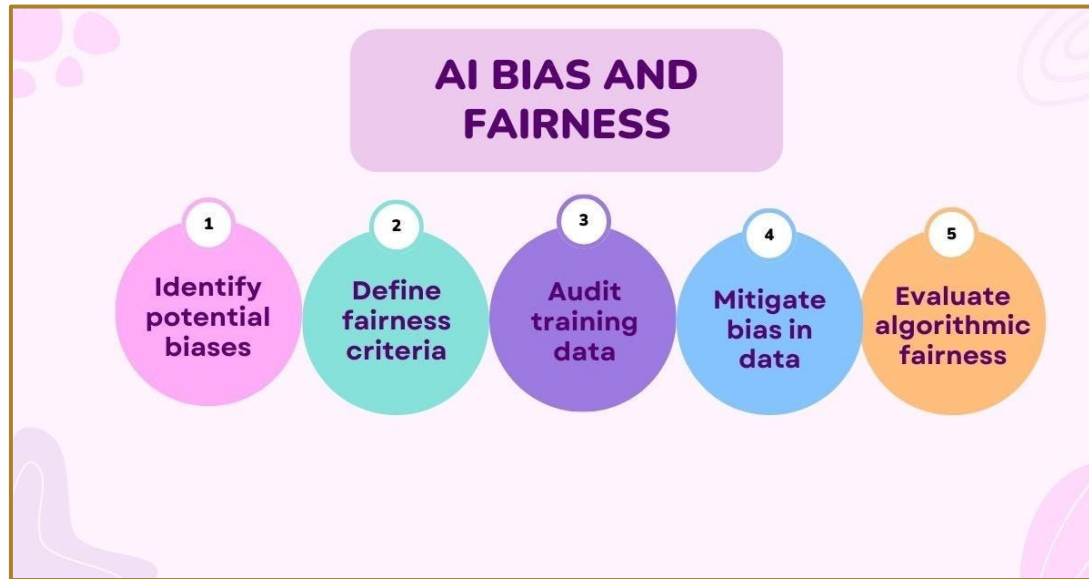
Principles of Ethical AI

Transparency: Can we explain how an AI system makes decisions? A black-box model with no interpretability poses ethical risks.



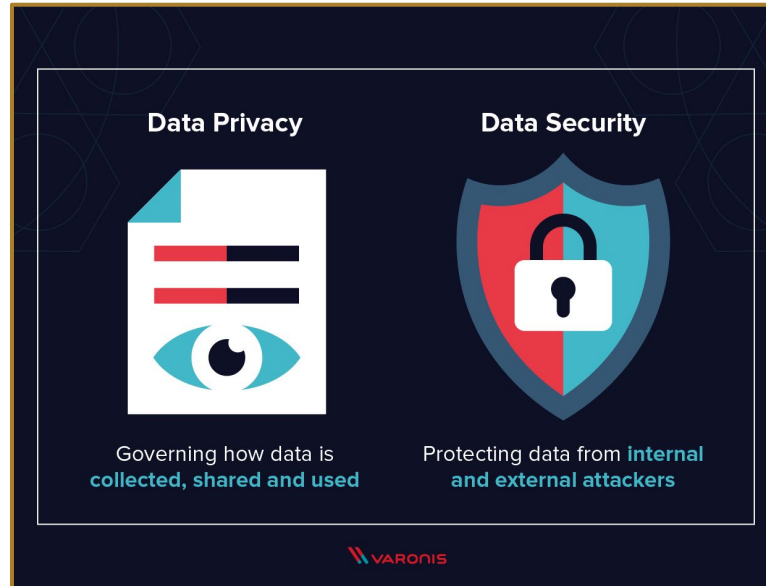
Principles of Ethical AI

Fairness: Are AI models treating all individuals equitably, or do they reinforce existing biases and discrimination?



Principles of Ethical AI

Privacy: Are users' data protected against misuse, and are they aware of how their data is used?



Example

- ❖ **Example:** Imagine an AI-powered hiring tool that screens job applications. If the algorithm unfairly prefers male candidates over equally qualified female candidates, ethical concerns about bias, accountability, and transparency arise.



Bias and Discrimination in AI

Bias and Discrimination in AI

- ❖ Bias in AI systems often stems from historical inequalities in training data, flawed assumptions in model design, or improper data processing. There are different types of biases in AI.

Types of Bias

- ❖ **Historical Bias:** When past discrimination is reflected in training data (e.g., biased hiring data favoring men leads to biased AI predictions).
- ❖ **Sampling Bias:** When a dataset underrepresents certain groups (e.g., facial recognition trained on lighter skin tones performs poorly on darker skin tones).
- ❖ **Algorithmic Bias:** When model design disproportionately favors certain groups (e.g., credit scoring models rejecting minorities due to systemic financial inequalities).

How to Identify and Mitigate Bias

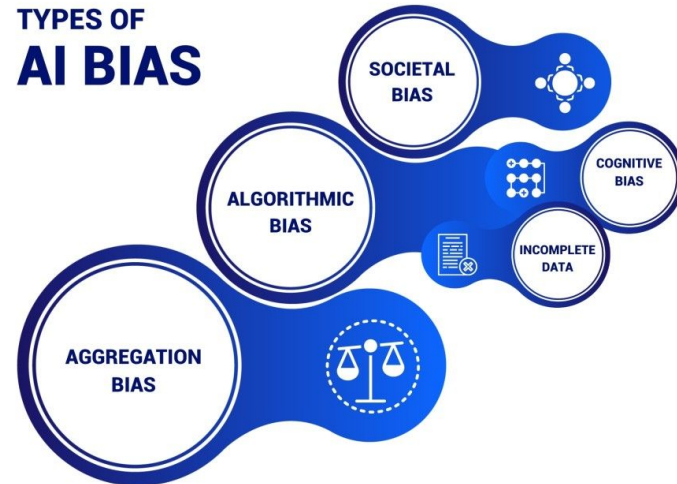
- ❖ **Dataset Auditing:** Analyze data distributions to detect skewed representation.
- ❖ **Bias Testing Metrics:** Use fairness metrics like disparate impact analysis.
- ❖ **Bias Mitigation Techniques:** Re-weighting training data, adversarial debiasing, and fairness constraints in algorithms.

How to Identify and Mitigate Bias

❖ **Analogy:**

- If you teach a child using biased history books, they will learn a skewed perspective of the world. Similarly, an AI system trained on biased data will produce biased outcomes.

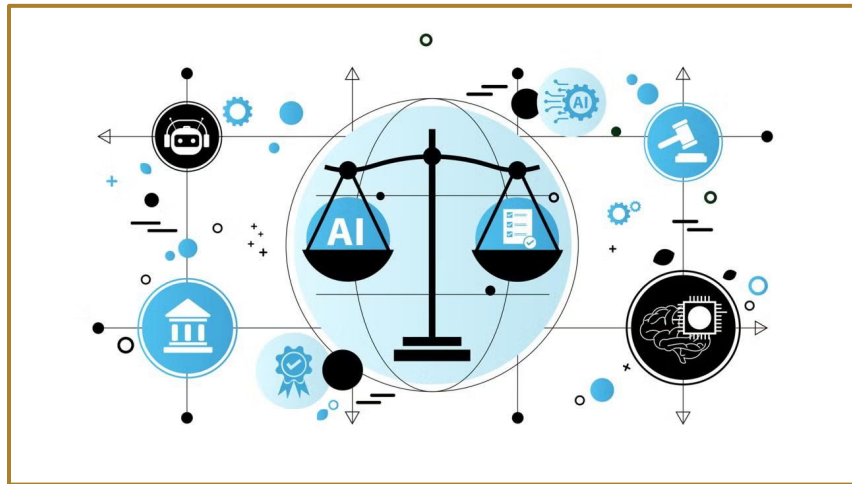
TYPES OF AI BIAS



Fairness in Algorithms

Fairness in Algorithms

- ❖ Ensuring fairness in machine learning models is a complex challenge, as fairness is a socio-technical issue with multiple definitions



Key Fairness Metrics

- ❖ **Demographic Parity:** Ensuring equal outcomes across groups.
- ❖ **Equalised Odds:** The model should have similar error rates for different demographic groups.
- ❖ **Fairness Through Awareness:** Ensuring that AI decisions do not rely on sensitive attributes like race or gender.



Techniques for Ensuring Fairness

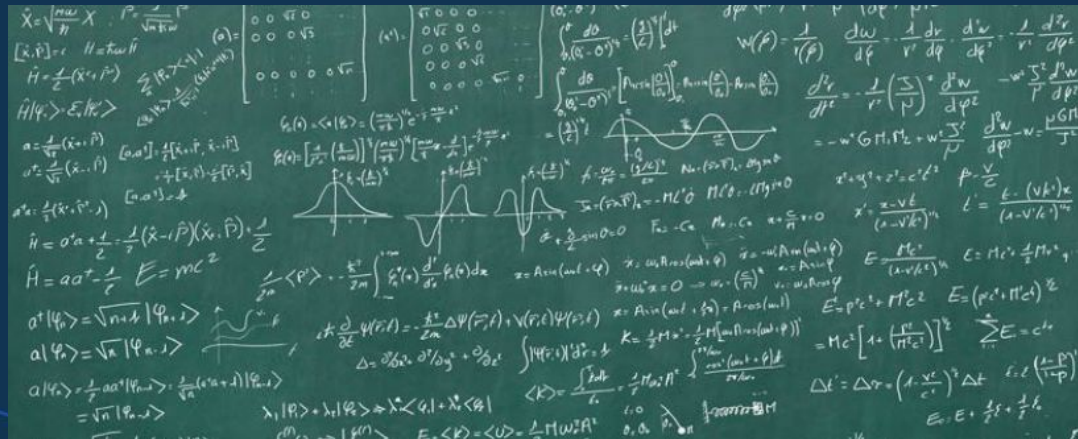
- ❖ **Pre-processing Methods:** Adjusting training data to balance representation.
- ❖ **In-processing Methods:** Modifying algorithms to incorporate fairness constraints.
- ❖ **Post-processing Methods:** Adjusting predictions to align with fairness goals.

Fairness

- ❖ **Example:** A predictive policing AI system disproportionately flags minority neighborhoods for crime prevention. Fairness-aware algorithms can help correct such biases.



Privacy Concerns in AI



Privacy Concerns in AI

- ❖ Privacy is a fundamental right, and AI systems must comply with data protection laws and ethical data handling practices.



Key Privacy Principles

- ❖ **User Consent:** Users should be aware of how their data is collected and used.
- ❖ **Minimization:** Collect only necessary data to achieve the intended purpose.
- ❖ **Anonymization:** Remove personally identifiable information to protect privacy.

Regulations & Ethical Guidelines

- ❖ **General Data Protection Regulation (GDPR):** Provides a legal framework for data privacy and user rights in the EU.
- ❖ **Differential Privacy:** Adds noise to data to prevent individual identification while maintaining dataset utility.



Regulations & Ethical Guidelines

◆ Example:

- A fitness app tracking users' locations must ensure data encryption and anonymization to prevent misuse of location data.



A woman with glasses and a patterned shirt is sitting at a desk, working on a computer. She is looking at a large monitor displaying code. A smartphone is on a stand next to her. The image has a teal overlay and a blue abstract shape on the left.

Accountability in AI

Explainable AI (XAI) and Responsible AI Development

- ❖ AI accountability ensures that decision-making processes are interpretable and traceable.
 - **Explainable AI (XAI):** Models that provide human-understandable reasons for decisions (e.g., SHAP values in machine learning).
 - **Model Auditing & Governance:** Third-party reviews to ensure AI compliance with ethical guidelines.
 - **Human-in-the-loop Systems:** Keeping human oversight in AI-driven decision-making.

Accountability in AI

❖ Example:

- If an AI medical diagnosis tool misclassifies a patient's condition, doctors must understand why the model failed and take corrective action.



Real-World Implications of Unethical AI

Case Studies

- ❖ **COMPAS Algorithm (Criminal Justice Bias):** The COMPAS risk assessment tool was found to disproportionately classify Black defendants as high risk for reoffending, showcasing racial bias in AI.
- ❖ **Cambridge Analytica (Data Privacy Violation):** Misuse of Facebook data for political manipulation highlighted the dangers of unethical data collection.
- ❖ **Amazon's Hiring AI (Gender Bias):** Amazon's hiring algorithm was found to favor male applicants due to biased historical data.

How to Prevent Ethical AI Failures

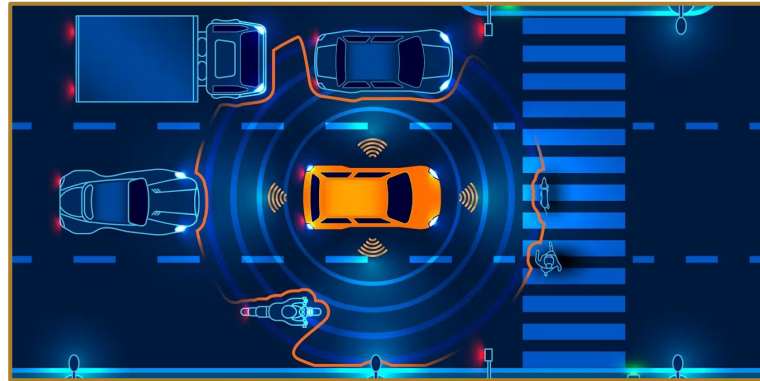
- ❖ Implement diverse and inclusive training datasets.
- ❖ Enforce AI ethics regulations and corporate accountability.
- ❖ Educate AI practitioners on ethical considerations.



Evaluation Metrics

❖ **Analogy:**

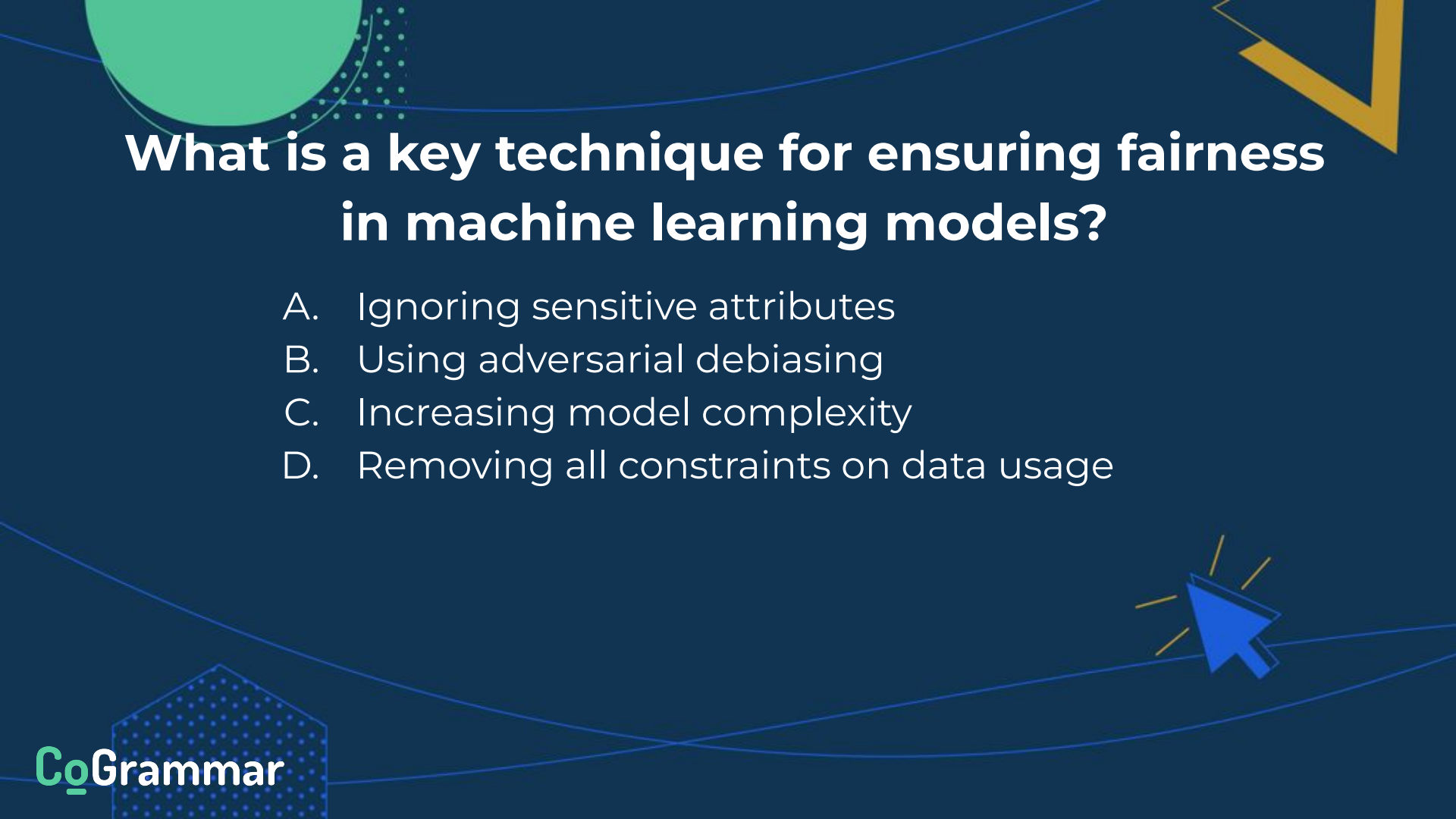
- Building an AI system without ethical considerations is like designing a self-driving car with no brake. Eventually, it will cause harm.



Conclusion: Building Ethical and Fair AI

- ❖ To ensure ethical AI, developers, organizations, and policymakers must work together to:
 - Prioritize fairness and transparency in AI systems.
 - Regularly audit AI models for bias and accountability.
 - Adhere to ethical frameworks and privacy regulations.
- ❖ As AI continues to evolve, ethical considerations must be at the forefront to ensure that technology serves humanity rather than harming it. Ethical AI isn't just an ideal, it's a necessity for a just and equitable society.

Polls Assessment



What is a key technique for ensuring fairness in machine learning models?

- A. Ignoring sensitive attributes
- B. Using adversarial debiasing
- C. Increasing model complexity
- D. Removing all constraints on data usage



Why is Explainable AI (XAI) important?

- A. It ensures algorithms run faster
- B. It helps stakeholders understand AI decisions
- C. It eliminates all bias in AI
- D. It prevents data breaches



What is one ethical concern related to AI and data privacy?

- A. Collecting user data without consent
- B. Ensuring model efficiency
- C. Using high-quality datasets
- D. Improving computational speed

Questions and Answers



Thank you for attending



CoGrammar



Department
for Education