HyperionDev

Welcome to this session:
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.

HyperionDev

What is Safeguarding?

Safeguarding refers to actions and measures aimed at protecting the human rights of adults, particularly vulnerable individuals, from abuse, neglect, and harm.

To report a safeguarding concern reach out to us via email:

safeguarding@hyperiondev.com

HyperionDev

Live Lecture Housekeeping:

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly.
- No question is daft or silly - ask them!
- For all non-academic questions, please submit a query: www.hyperiondev.com/support
- To report a safeguarding concern reach out to us via email: safeguarding@hyperiondev.com
- If you are hearing impaired, please kindly use your computer's function through Google chrome to enable captions.

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.**

Relevance

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?

HyperionDev

Relevance

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?*

HyperionDev

Relevance

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.

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?

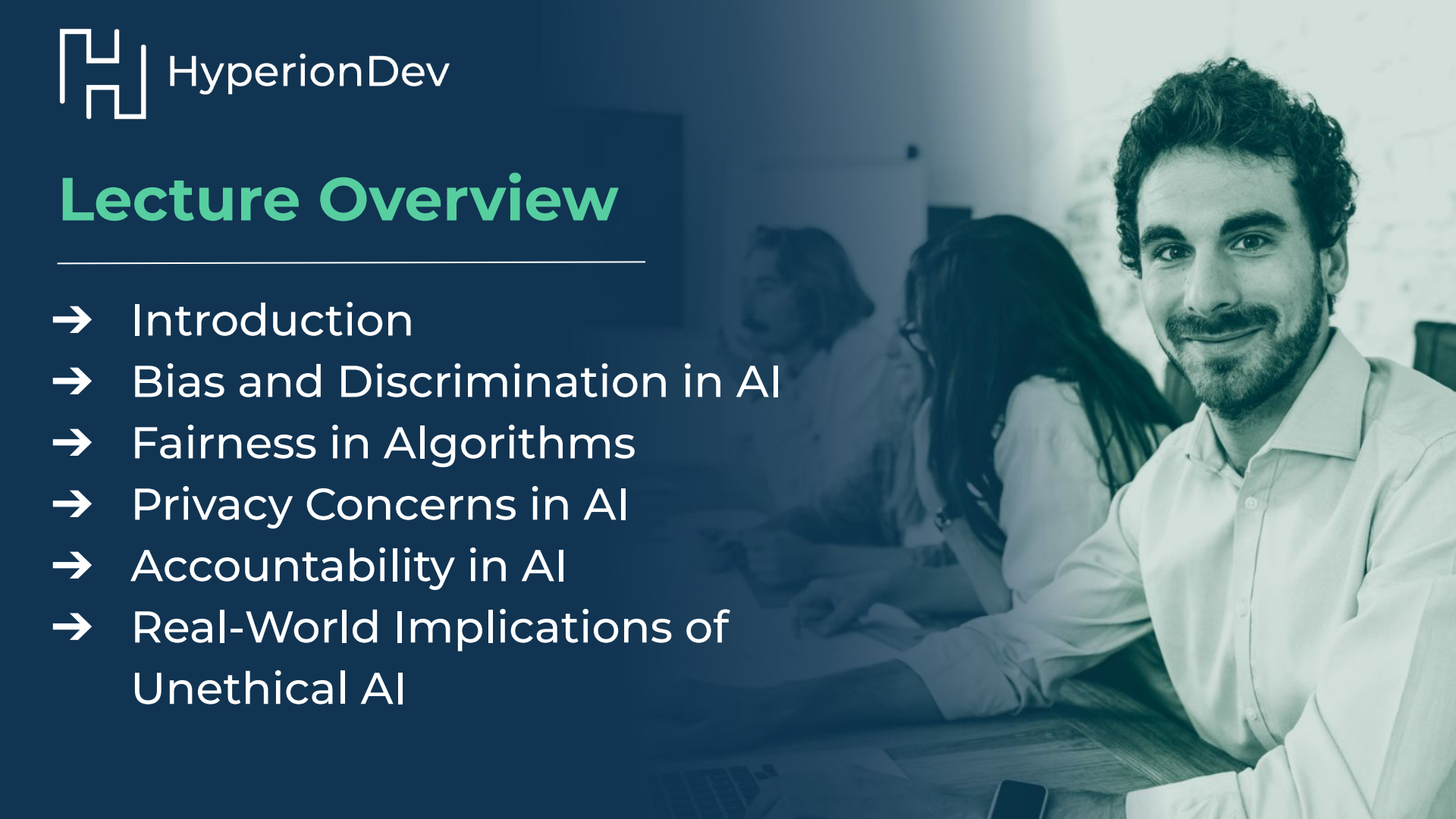
A. A model favouring one demographic group over another
B. A model providing consistent and fair results for all users
C. An algorithm that adapts to user preferences
D. A model that ignores sensitive attributes

HyperionDev

HyperionDev

Lecture Overview

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



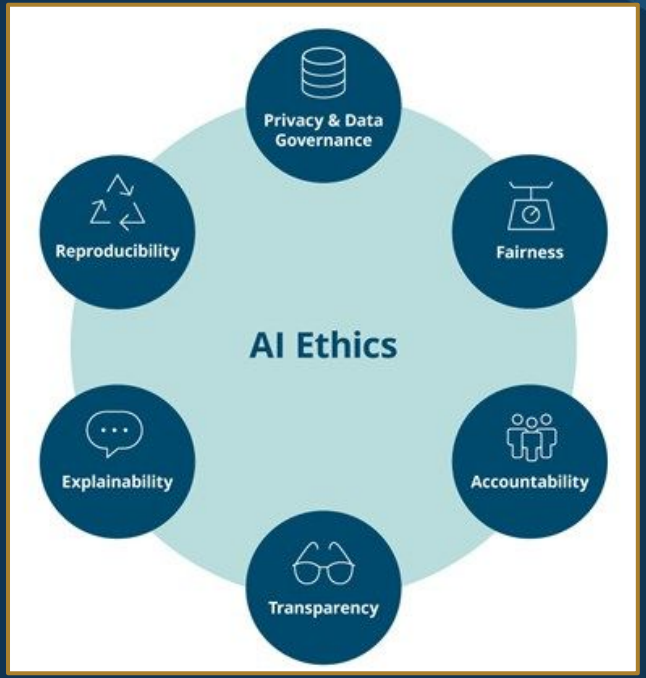
Introduction to Ethics and Fairness in AI and Data Science

HyperionDev

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.

Image source: LinkedIn



HyperionDev

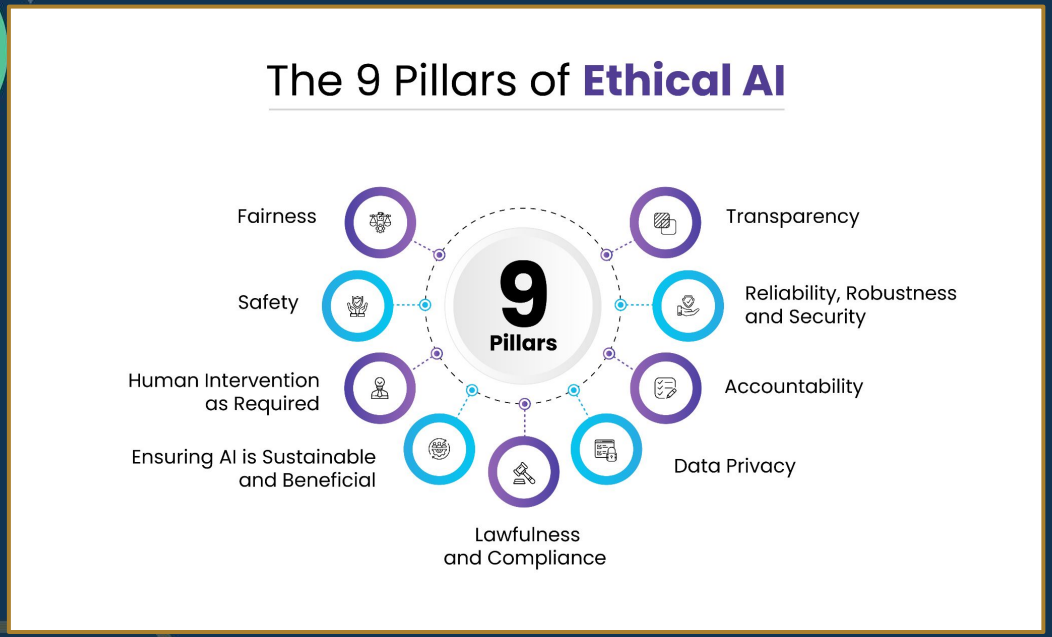
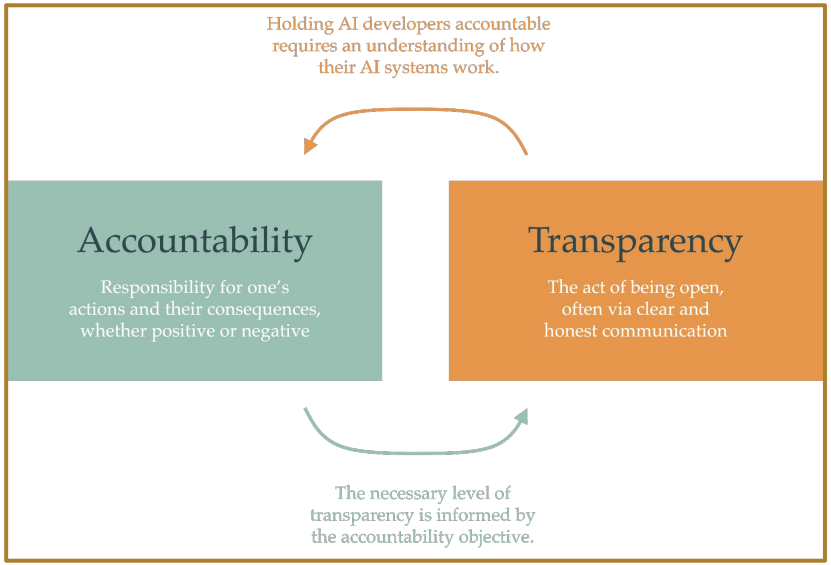


Image source:
https://www.cogentinfo.com/resources/9-ethical-ai-principles-for-organizations-to-follow

Ethical AI is built upon key principles

HyperionDev

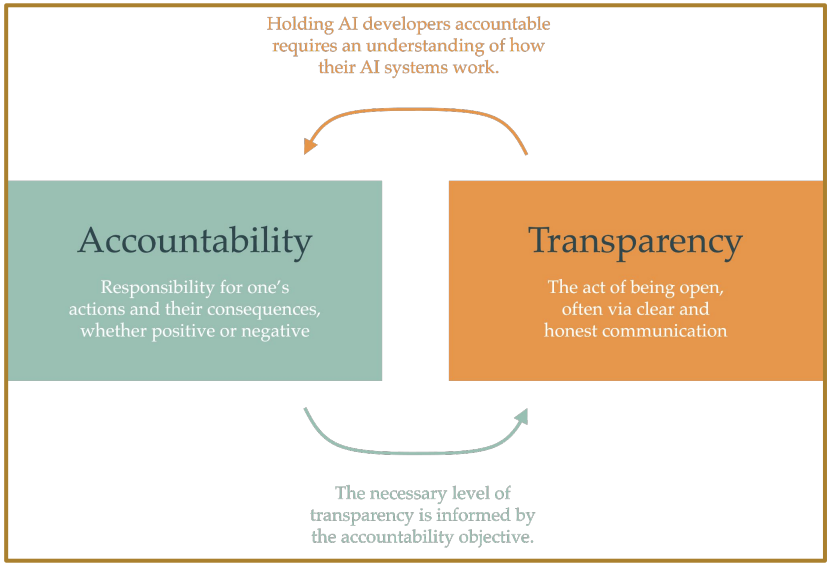
Principles of Ethical AI



- **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?

HyperionDev

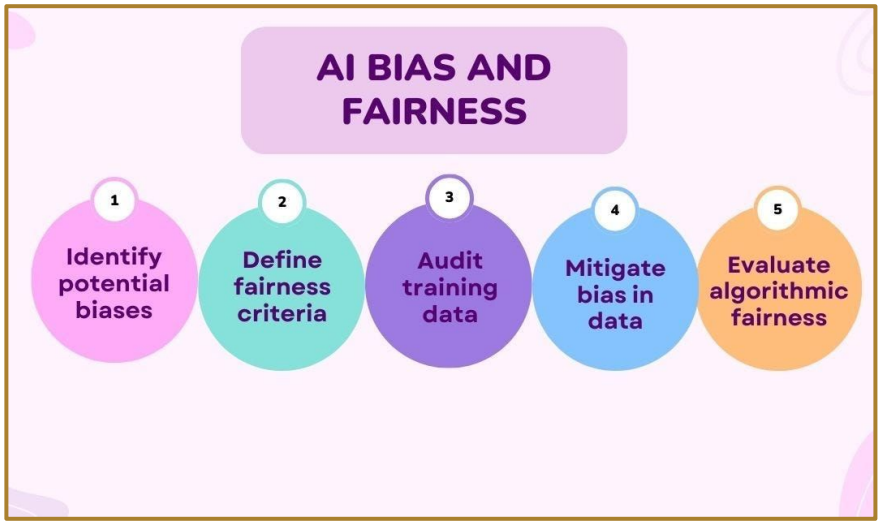
Principles of Ethical AI



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

HyperionDev

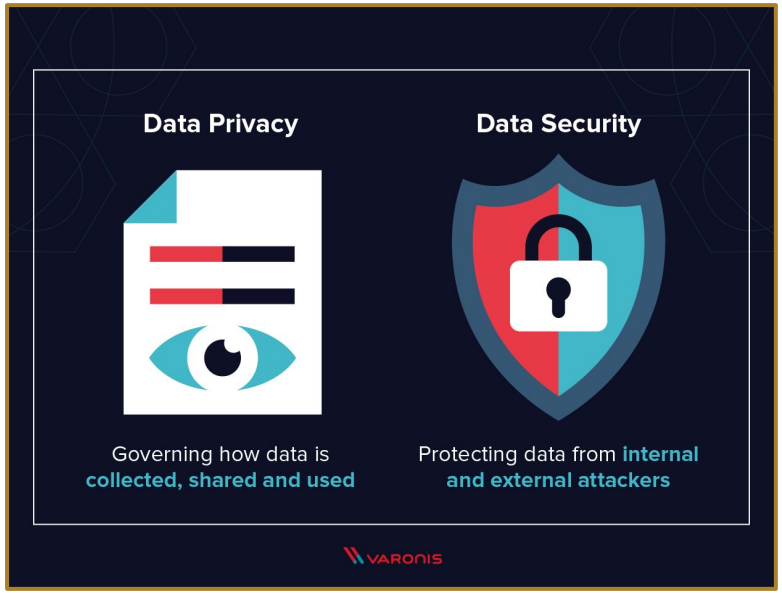
Principles of Ethical AI



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

HyperionDev

Principles of Ethical AI

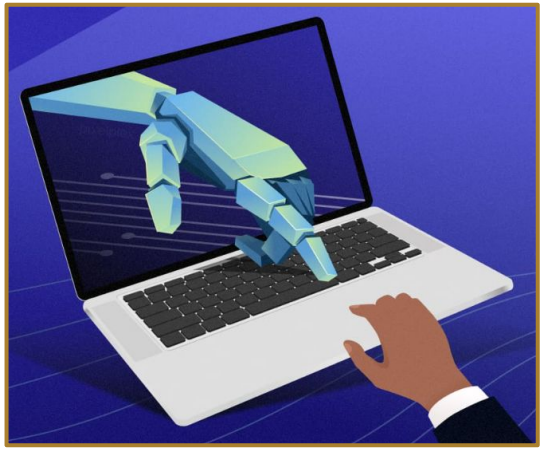


- **Privacy:** Are users' data protected against misuse, and are they aware of how their data is used?

HyperionDev

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.



HyperionDev

Bias and Discrimination in AI

HyperionDev

Bias and Discrimination in AI

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

Types of Bias

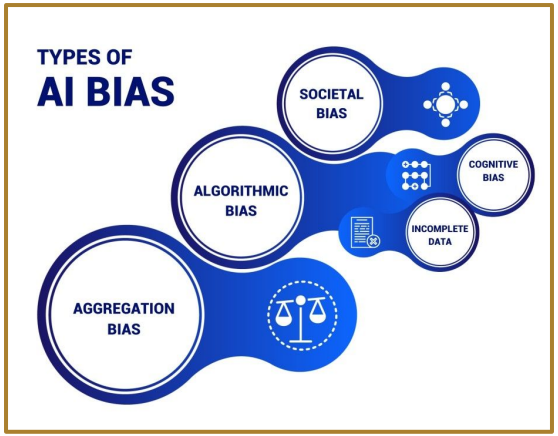
- **Historical Bias:** When past discrimination is reflected in training data (e.g., biased hiring data favoring men leads to biased AI predictions).
- **Sampling Bias:** When a dataset underrepresents certain groups (e.g., facial recognition trained on lighter skin tones performs poorly on darker skin 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

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.



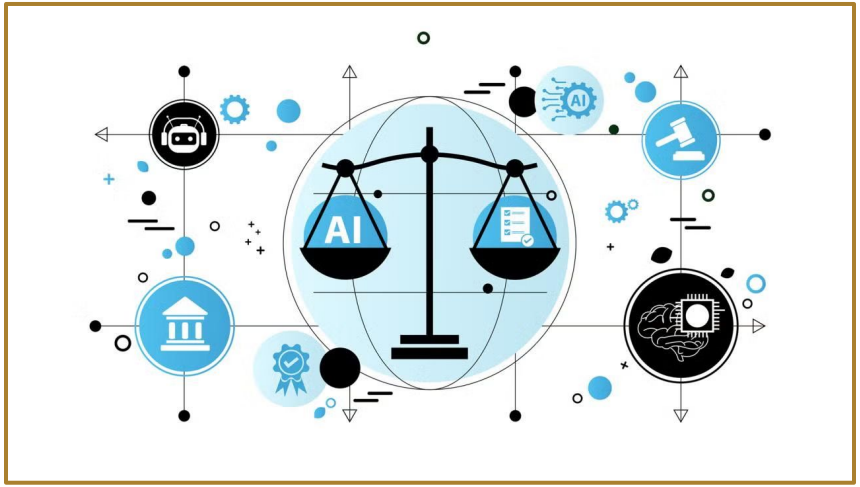
HyperionDev

Fairness in Algorithms

HyperionDev

Fairness in Algorithms

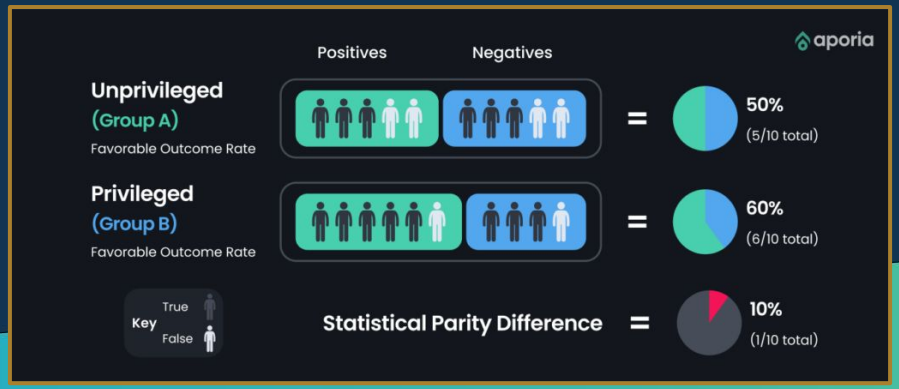
Ensuring fairness in machine learning models is a complex challenge, as fairness is a socio-technical issue with multiple definitions



HyperionDev

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.



HyperionDev

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.

HyperionDev

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.

HyperionDev

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.

HyperionDev

Example

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

HyperionDev

BREAK

HyperionDev

Accountability in AI

HyperionDev

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.

Example

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

HyperionDev

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.

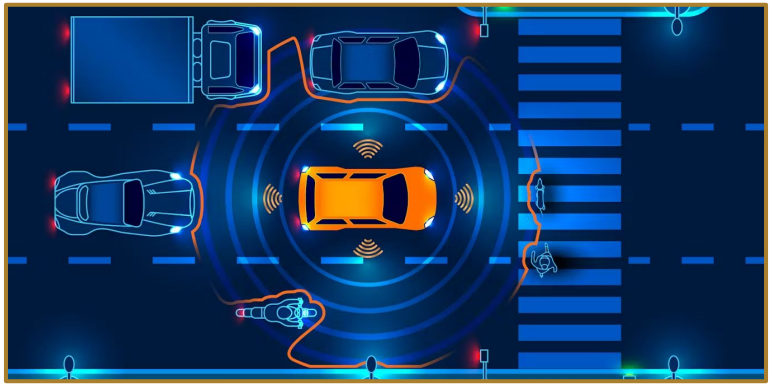
How to Prevent Ethical AI Failures

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

HyperionDev

Ethical Considerations

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



HyperionDev

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.

HyperionDev

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

HyperionDev

Q & A SECTION

Please use this time to ask any questions relating to the topic, should you have any.


Thank you
for attending
HyperionDev