

ETHICAL AI ASSESSMENT

→ Presented by DevTeam

OVERVIEW

Theoretical Understanding

01

Case Study Analysis

02

Practical Dataset Audit

03

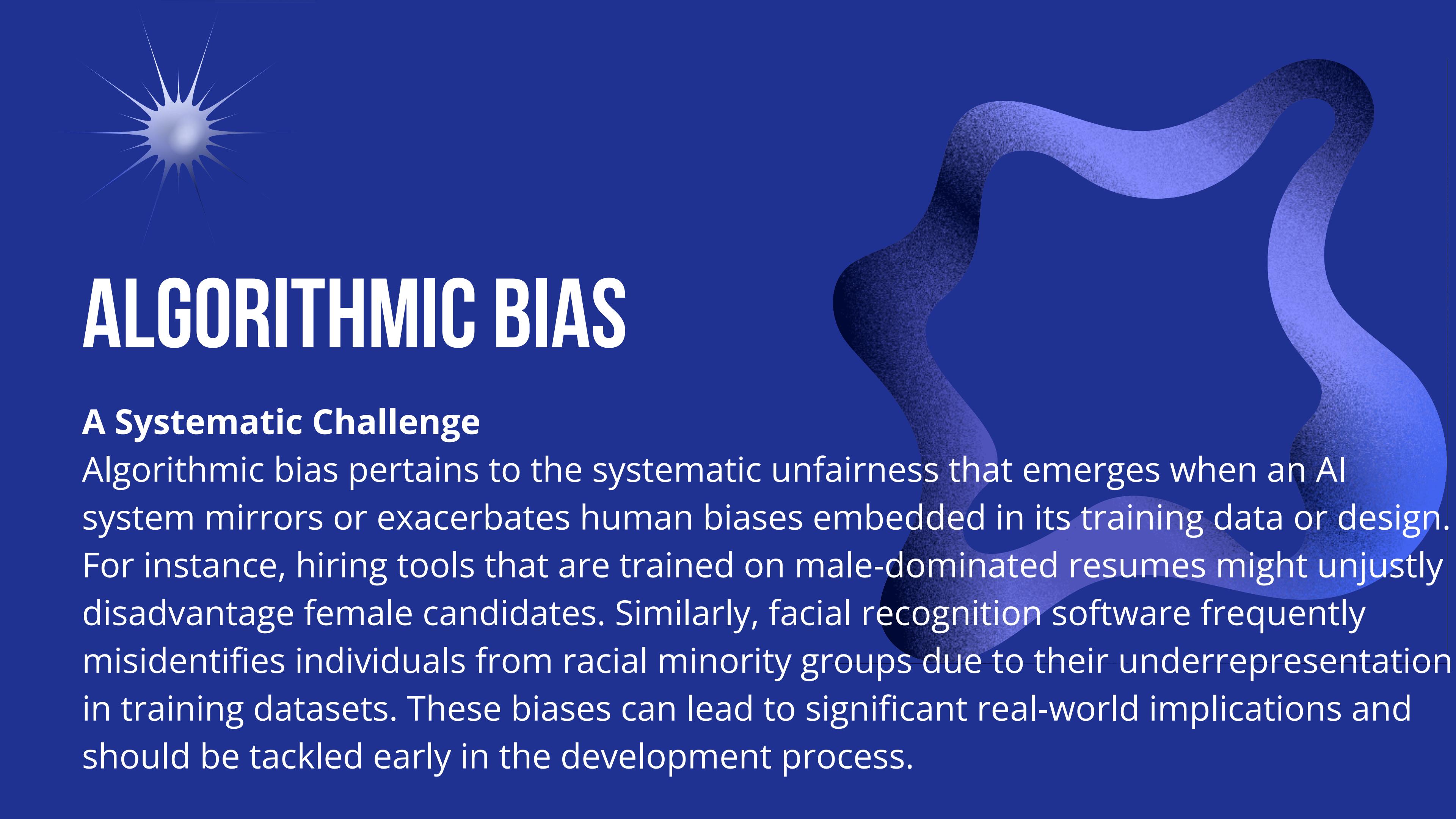
Personal Ethical Reflection

04

Ethical Policy for Healthcare AI

05

Conclusion



ALGORITHMIC BIAS

A Systematic Challenge

Algorithmic bias pertains to the systematic unfairness that emerges when an AI system mirrors or exacerbates human biases embedded in its training data or design. For instance, hiring tools that are trained on male-dominated resumes might unjustly disadvantage female candidates. Similarly, facial recognition software frequently misidentifies individuals from racial minority groups due to their underrepresentation in training datasets. These biases can lead to significant real-world implications and should be tackled early in the development process.

TRANSPARENCY VS EXPLAINABILITY



Transparency

Clear processes ensure trust in AI systems.



Explainability

concerns how well the AI's decisions can be understood by humans. Both are critical in fostering trust



Fairness

Equitable treatment is essential for all users.

GDPR'S IMPACT ON AI

The General Data Protection Regulation (GDPR) directly shapes how AI is developed and used within the European Union. It mandates that users must give explicit consent before their data is processed by an algorithm, the right to an explanation when automated decisions impact individuals. GDPR also restricts profiling and requires developers to implement protective measures such as data minimization and encryption. These rules promote the adoption of ethical and legally sound practices.

BIASED HIRING TOOL — AMAZON

Amazon's recruiting algorithm favored male candidates, replicating bias found in historical hiring data. This occurred because the system was trained primarily on resumes from previously successful applicants—most of whom were men. To correct this, developers could rebalance training data, filter out gender-correlated features, and use fairness-aware models. Post-correction evaluation should rely on metrics like statistical parity and equal opportunity difference to ensure improvements are equitable.

FACIAL RECOGNITION IN POLICING

Facial recognition tools often misidentify minorities, leading to ethical issues such as wrongful arrests and privacy violations. To responsibly use these technologies, strict limitations, third-party bias audits, and community oversight are essential to protect civil liberties and prevent systemic discrimination.

INTRODUCTION TO ETHICAL AI

This section will explore the importance of ethics in artificial intelligence, highlighting key principles and considerations that guide the responsible development and deployment of AI technologies.



COMPAS DATASET BIAS AUDIT SUMMARY

Using the IBM AI Fairness 360 toolkit, an audit of the COMPAS recidivism prediction dataset revealed racial bias in its risk scores. Black defendants were flagged as high-risk significantly more often than White defendants, despite similar circumstances. Metrics like disparate impact and positive rate difference quantified these disparities, underscoring the need for more equitable modeling practices in the justice system.

BIAS REMEDIATION & ETHICAL IMPACT

After identifying unfairness in the COMPAS dataset, bias mitigation techniques such as Equalized Odds post-processing were used to rebalance predictions. These adjustments reduced group disparities while preserving model performance. The audit revealed the importance of auditing feature selections, applying fairness constraints throughout development, and engaging stakeholders in remediation. Ethical AI isn't just a technical fix; it's a commitment to a just philosophy.

PERSONAL ETHICAL REFLECTION

In a previous academic project predicting student dropout risk, I faced ethical dilemmas around privacy and bias. If I revisited that project today, I would insist on explicit consent from participants, implement fairness audits, and ensure all predictions were explainable to educators and students. Human oversight would be central to all decisions, keeping the system aligned with the ethical principles of autonomy, justice, and non-maleficence.



ETHICAL AI IN HEALTHCARE

Responsible healthcare AI begins with informed patient consent and strong data protection measures. Bias mitigation must be prioritized through diverse data representation and continual fairness checks. Transparency is also vital for trust between clinicians and patients. Human oversight should remain active for a balance of accountability and innovation, supported by independent audits and ethical review boards that hold systems accountable for patient welfare.

CONCLUSION & FINAL THOUGHTS

Ethical AI design is essential, not just for compliance but to protect human dignity. By grounding systems in fairness, transparency, and empathy, developers can prevent harm and ensure AI benefits are equitably distributed. As AI's reach continues to grow, so does our responsibility to create systems that reflect our highest values. This presentation is both a blueprint and a call to action: build AI that respects, uplifts, and protects.