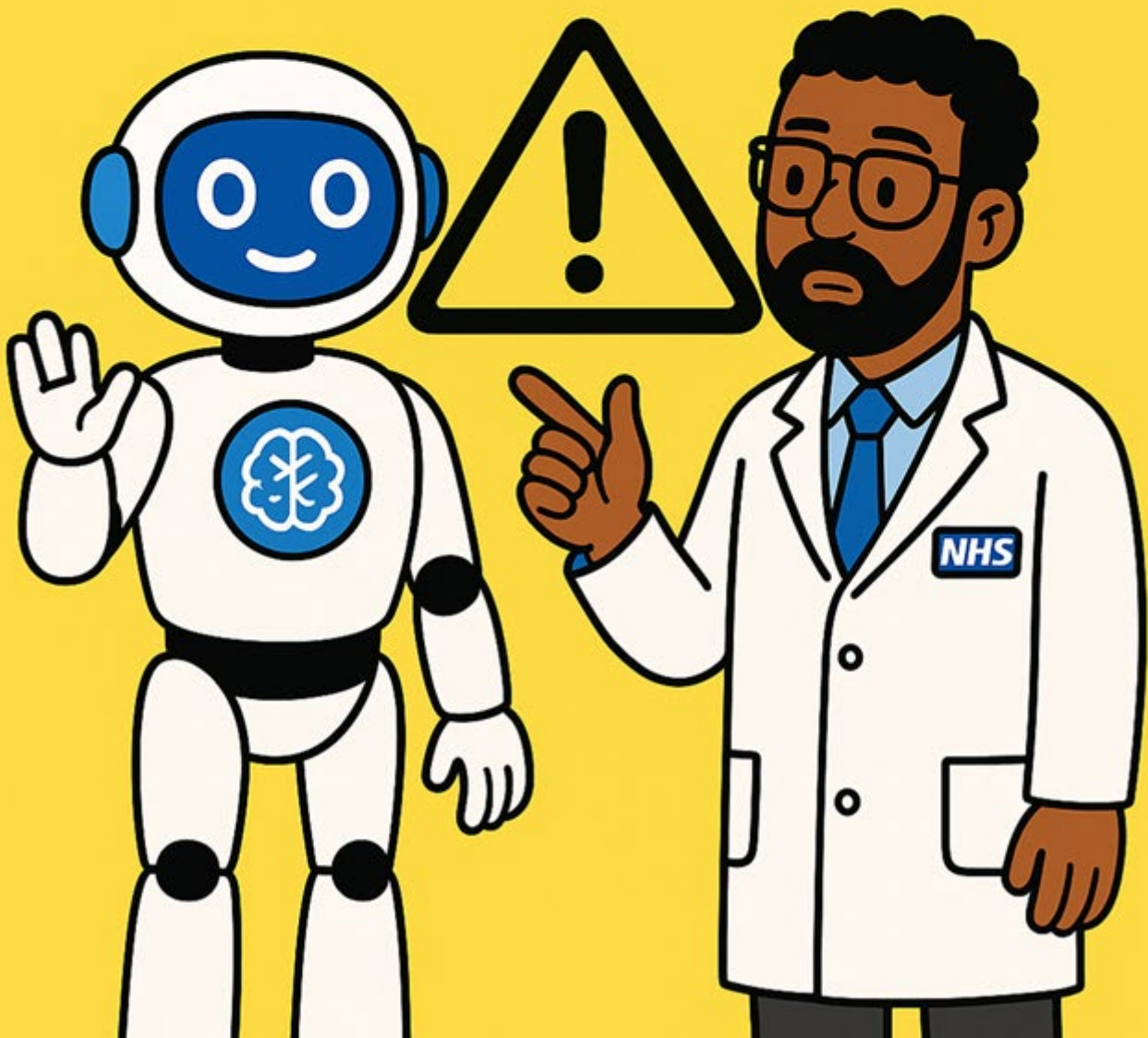


Simple AI

Risk and Bias



What Is AI Bias?

Explaining AI bias in Plain English

AI bias happens when a computer system gives results that are unfair or unbalanced. This usually comes from the **data used** to train the system. If the data reflects inequalities, the AI can repeat them—or make them worse.



Mini examples



Unequal outcomes for patients



Bias in staff hiring

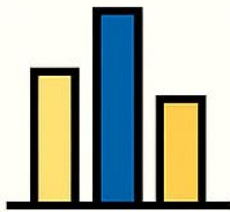


Skewed reporting

Where Does Bias Come From?

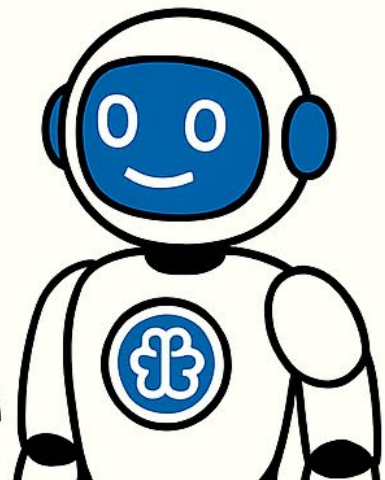
Common sources of AI bias explained simply

AI bias can creep in at many stages. It's not always intentional — but it can still cause harm if left unchecked.



Data Bias

Training data doesn't represent everyone fairly



Algorithm Bias

The model itself gives skewed results



Human Bias

Designers' own assumptions



Feedback Bias

Past errors get reinforced

Why Bias Matters in the NHS

Real risks for patients, staff, and services

Bias in AI isn't just technical – it has real-world impacts on healthcare. If left unchecked, it could affect safety, fairness, and trust in the NHS.



Patient Safety

Biased tools may miss conditions in underrepresented groups



Workforce Decisions

Recruitment and rostering could disadvantage staff



Audit and Assurance

Skewed results undermine evidence and oversight



Public Trust

If people think AI is unfair, NHS innovation will lose support

What Is Algorithmic Fairness?

Making sure AI treats people equally

Algorithmic fairness is about checking that AI systems don't favour one group over another. In the NHS, fairness means every patient and staff member is treated with equal care and accuracy.



Equal Opportunity



Outcomes should not depend on age, gender, ethnicity, or postcode



Fairness Metrics

Involving diverse voices makes systems fairer

Audit and Oversight



Regular reviews catch unfair patterns early

Why Explainability Matters

Understanding how AI makes decisions

Explainability means being able to understand how and why an AI system reached its result. In healthcare, decisions must be transparent so patients, staff, and auditors can trust them.



Clinician Support

Clinicians need to know why AI suggests a diagnosis



Fairness Metrics

Clear evidence for assurance and accountability



Challenge Decisions

Explainable AI allows humans to question errors

Black-box AI is risky – explainability keeps NHS decisions safe and accountable.

Managing the Risks

NHS teams can reduce bias risks by:

- Conducting Data Protection Impact Assessments (DPIAs).
- Applying NHS AI Assurance and DTAC guidance.
- Testing AI across diverse patient groups.



Embedding continuous monitoring, not just one-off checks.

What to Look Out For

Warning signs of bias in AI:

Results that differ systematically between groups.

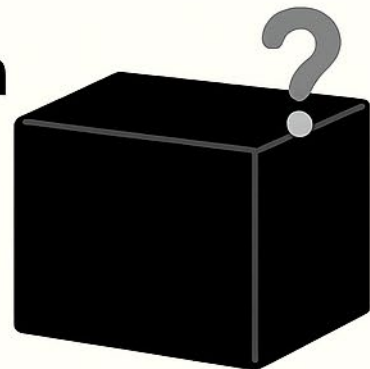


Over-reliance on historical data that reflects old inequalities.



Lack of transparency in how results are produced.

'Black box' systems with no clear explanations



Good Practice Checklist

- ☒ Test datasets for balance and representation
- ☒ Document how the algorithm works
- ☒ Involve clinicians, patients, and data experts
- ☒ Regularly audit outcomes for fairness
- ☒ Escalate risks where bias is suspected
- ☒ Involve clinicians, patients and data experts

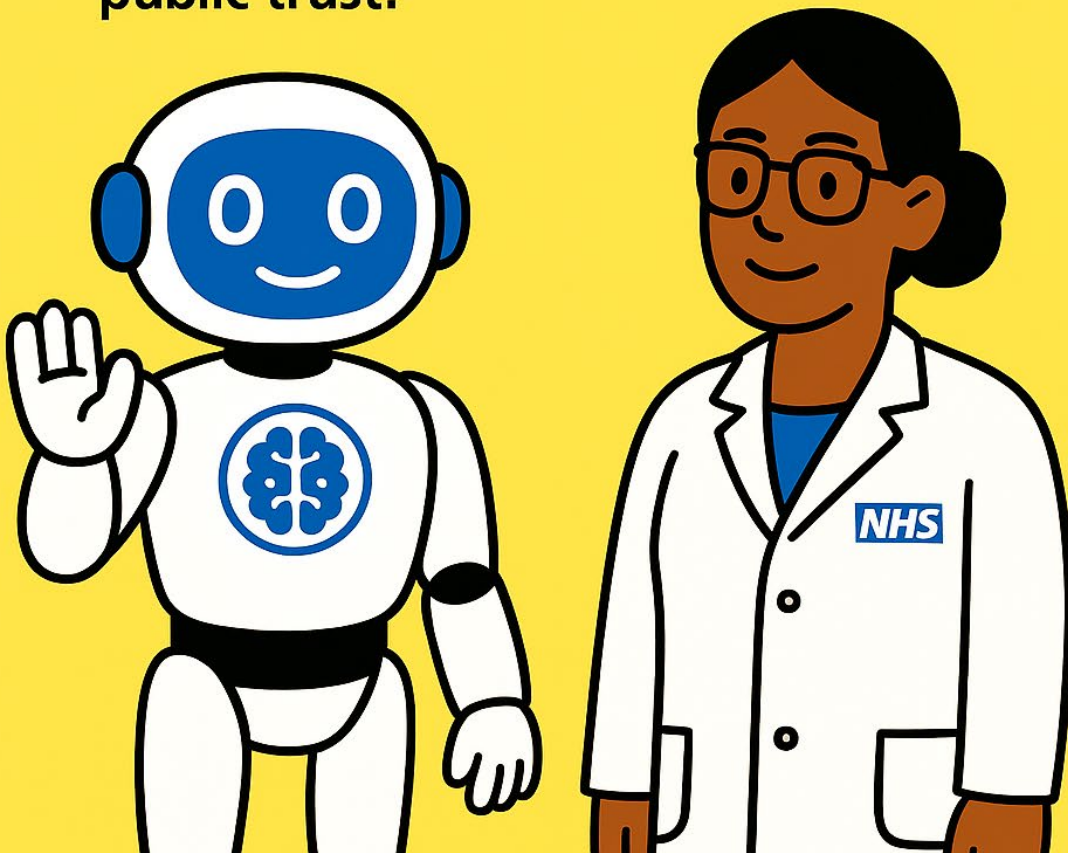




Final Thoughts

AI brings great opportunities for the NHS – but fairness must be built in from the start.

Understanding bias, testing for it, and explaining decisions clearly will help protect patients, staff, and public trust.



Need help? MIAA can support your next move.

Email: MIAA.Admin@miaa.nhs.uk **Tel:** +44 (0)151 285 4500

© MIAA, May 2025