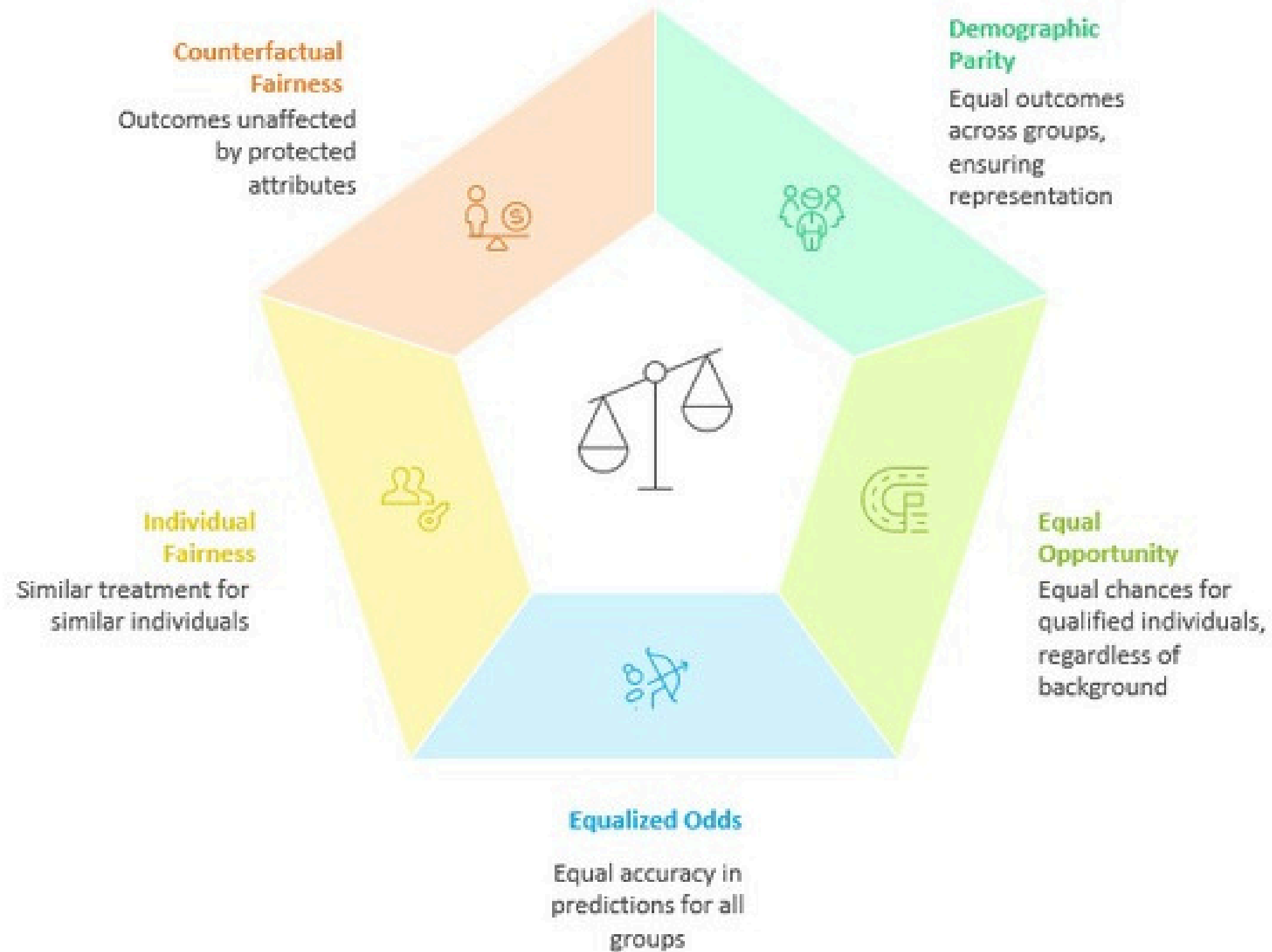# Responsible AI Starts with Ethics: A Step-by-Step Primer
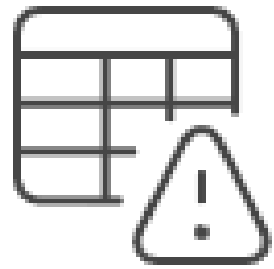
Fairness is one of the core Ethical principles of AI

What is Fairness means?

Fairness means treating individuals and groups equitably when algorithms make decisions that affect their lives—like getting a loan, a job interview, or healthcare.

Types of Fairness in AI

Counterfactual Fairness — Outcomes unaffected by protected attributes

Demographic Parity — Equal outcomes across groups, ensuring representation

Individual Fairness — Similar treatment for similar individuals

Equal Opportunity — Equal chances for qualified individuals, regardless of background

Equalized Odds — Equal accuracy in predictions for all groups

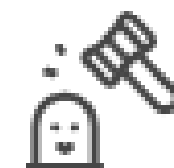**How to identify and address bias in training data?**

Historical Bias

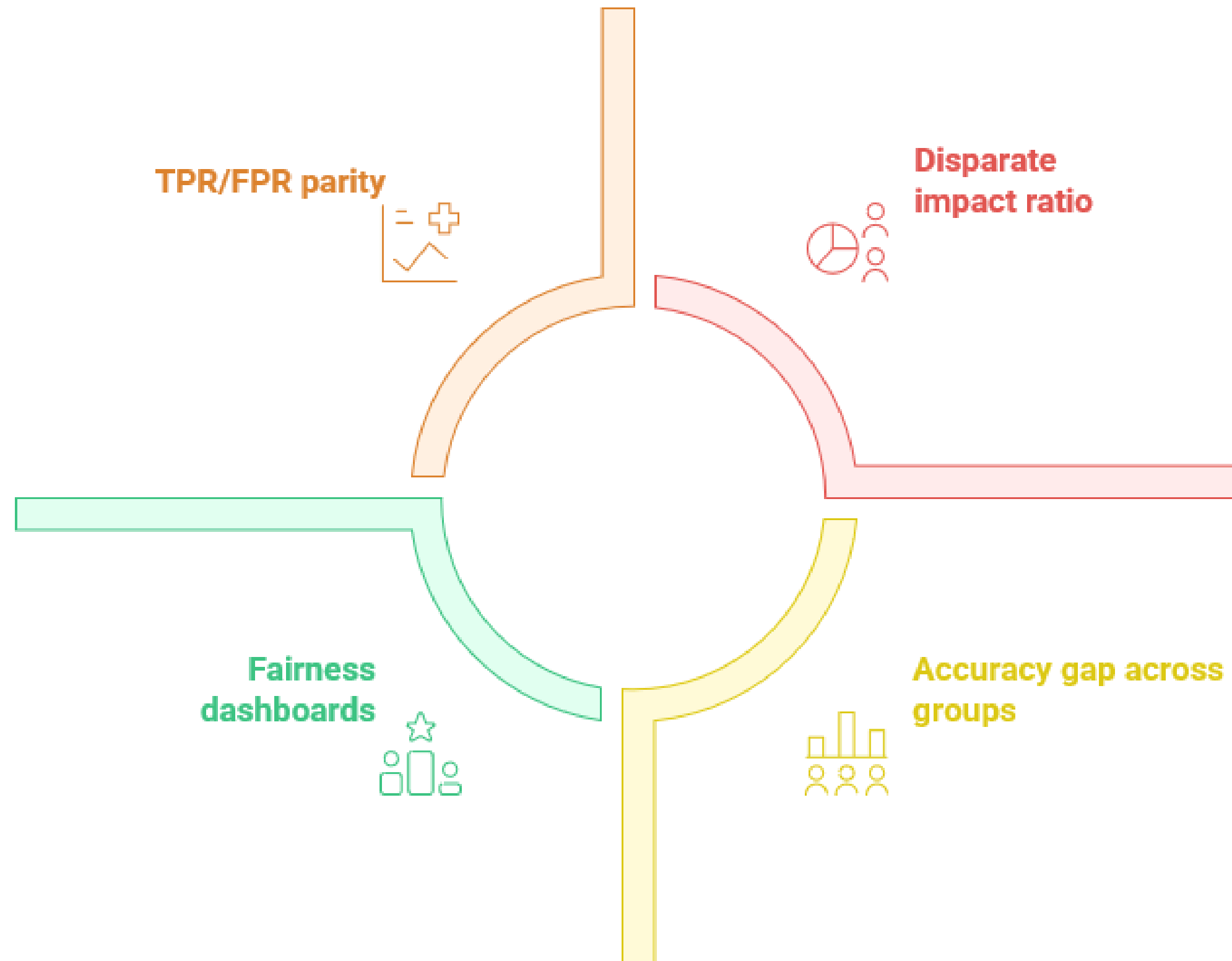Reflects past unfairness in data, like gender bias in hiring records.

Sampling Bias

Underrepresents certain groups, such as rural patients in medical data.

Label Bias

Labels reflect judgment, not truth, like police records indicating arrests, not crime.

Measure and Monitor Fairness

TPR/FPR parity

Disparate impact ratio

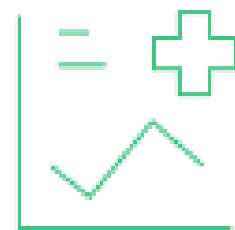Fairness dashboards

Accuracy gap across groups
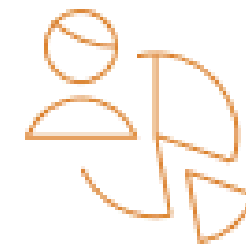
How to ensure fairness?

Consult Experts

Gain insights from domain experts, ethicists, and users to understand diverse perspectives.

Consider Context

Recognize that fairness varies across sectors like healthcare and education.

Evaluate Trade-offs

Distinguish between legal, moral, and business considerations.

Address Social Determinants

Ensure interventions are just by considering social factors.
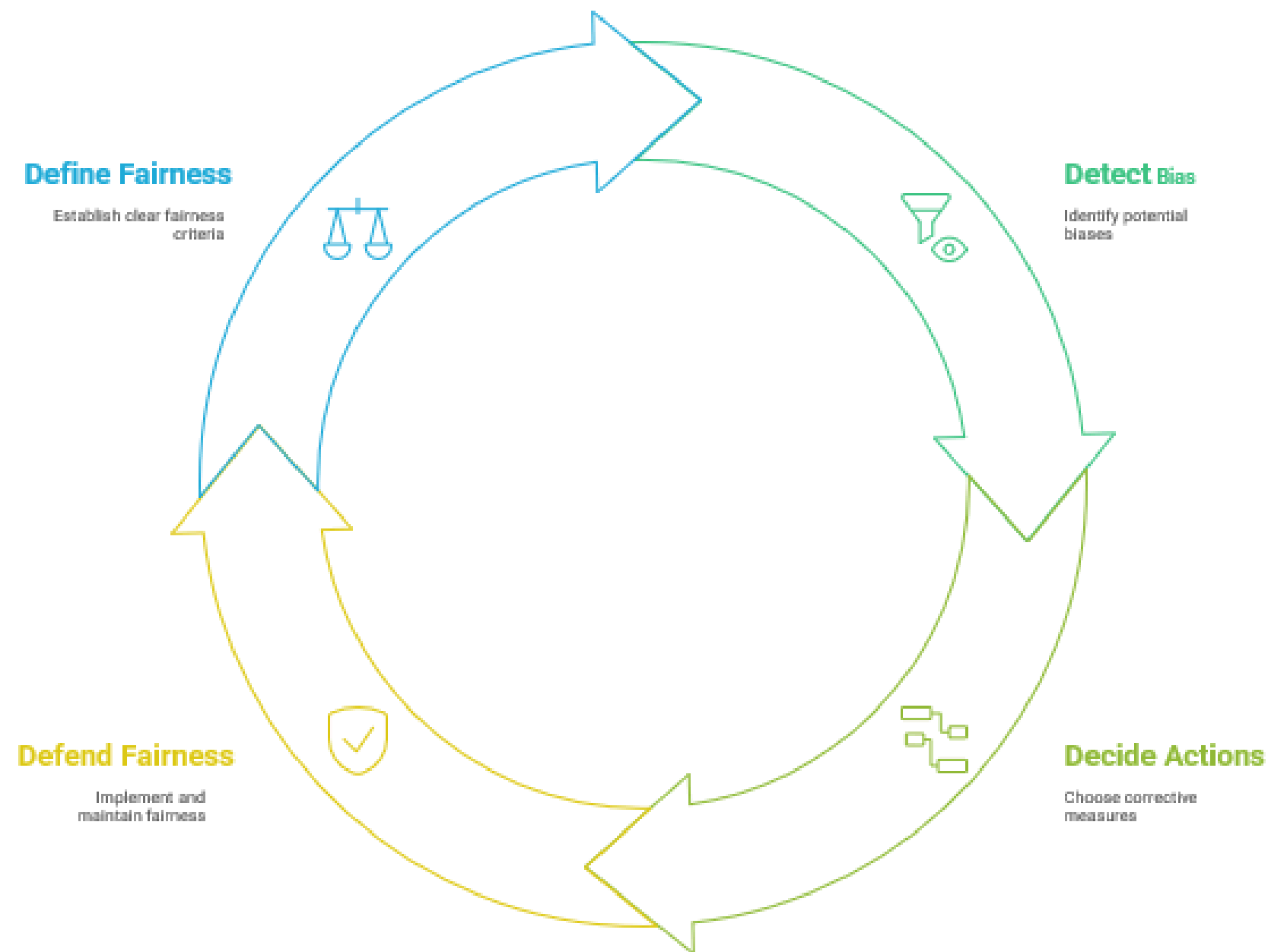
AI Documentation Standards

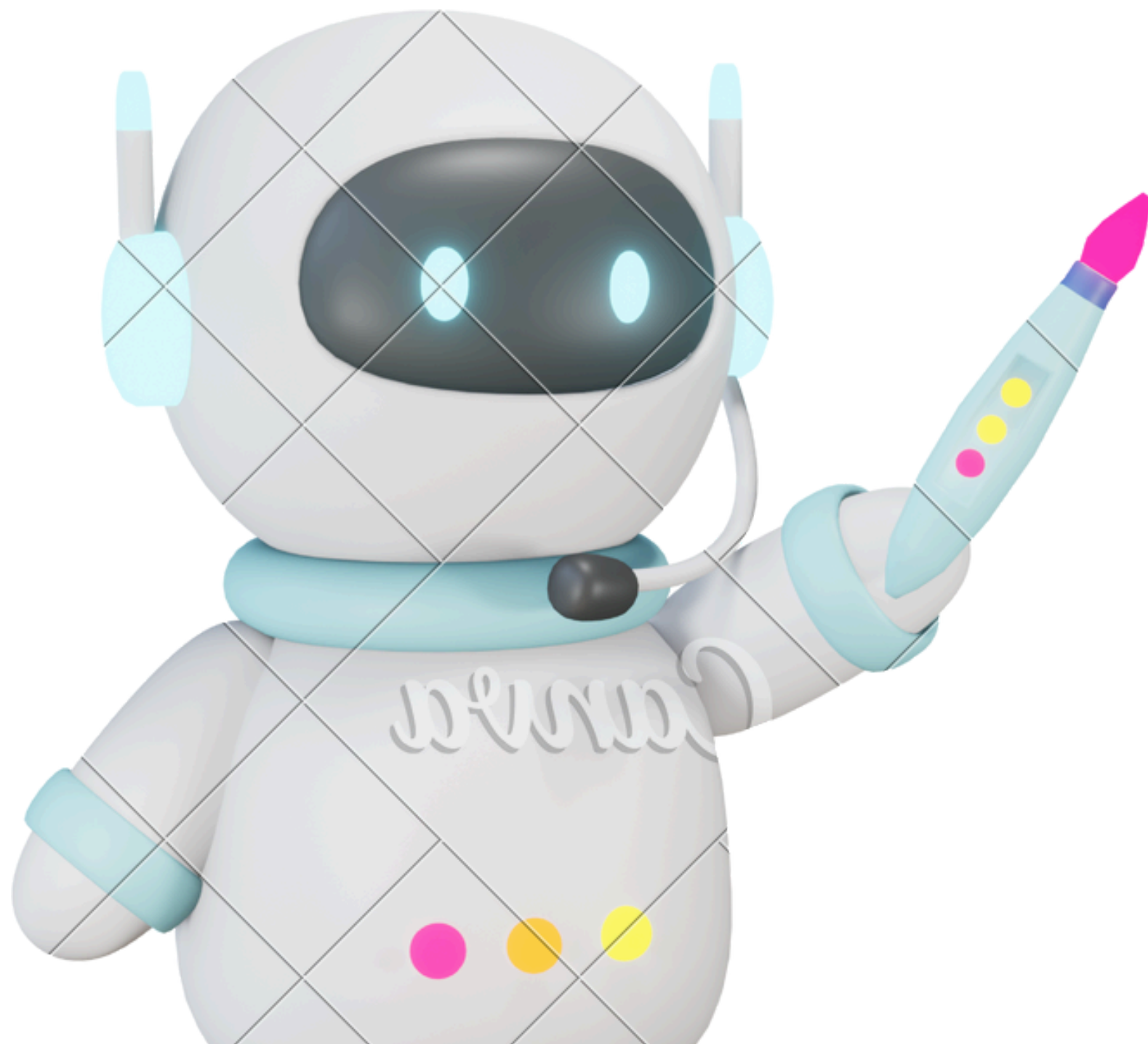| Format	Focus	Used By	Related To
Datasheets	Datasets	Microsoft, Meta, academia	Standard benchmark
Data Cards	Datasets (user-focused)	Google	Usability-enhanced
FactSheets	Services & datasets	IBM	Enterprise trust
Model Cards	Models	Hugging Face, Google	Transparency
System Cards	Full AI systems	OpenAI, DeepMind, Anthropic	Safety & policy
Data Statements (NLP)	Linguistic datasets	Academia, Google	NLP-specific
Nutrition Labels	Datasets	MIT, IBM	Consumer transparency
AI Impact Assessments	Policy + Governance	Governments, NGOs, regulators	Legal accountability

Always document
- What fairness definition used and why
- What metrics you tracked
- What trade-offs you made

Building Responsible Systems
is a
Continuous Commitment

Define Fairness

Establish clear fairness criteria

Detect Bias

Identify potential biases

Defend Fairness

Implement and maintain fairness

Decide Actions

Choose corrective measures

Thank You

Build powerful AI.
Build with ethics.
Build together.