Bias ist nicht gleich Bias:
Fünf Fehlerquellen bei KI-Entscheidungen

Eine gute Modellkennzahl kann eine schlechte Entscheidung stützen. Entscheidend ist, wo die Grundlage nicht trägt – und welche Gegenprüfung das klärt.

FÜNF FEHLERBEREICHE – FÜNF UNTERSCHIEDLICHE GEGENPRÜFUNGEN

1 Menschliches Urteil & Gruppendynamik

TYPISCHER MECHANISMUS
Confirmation Bias, Anchoring, einseitiger Informationsaustausch

GEGENPRÜFUNG
Gegenbelege und besondere Fachinformationen getrennt einholen; klären, welcher Befund die Empfehlung ändern würde.

2 Daten, Messung & Analyse

TYPISCHER MECHANISMUS
Selection Bias, unklare Referenzbewertungen, Data Leakage

GEGENPRÜFUNG
Fehlende Produktionsvarianten prüfen, Fehlerkriterien vereinheitlichen und Trennung von Trainings- und Testdaten nachvollziehen.

3 Modell, Evaluation & Anwendung

TYPISCHER MECHANISMUS
Distribution Shift, ungeeignete oder isolierte Kennzahlen

GEGENPRÜFUNG
Sensitivität, übersehene Fehler und Fehlalarme unter relevanten Einsatzbedingungen gemeinsam bewerten.

4 Projektplanung & Wirtschaftlichkeit

TYPISCHER MECHANISMUS
Optimism Bias, Planning Fallacy

GEGENPRÜFUNG
Tatsächlich veränderbare Prüfabläufe und Kapazitäten ermitteln; Aufwand sowie manuelle und regelbasierte Alternativen vergleichen.

5 Organisation, Governance & Betrieb

TYPISCHER MECHANISMUS
Unklare Zuständigkeiten und fehlende Kontrolle – keine Biases im engeren Sinn

GEGENPRÜFUNG
Freigaben, Eingriffsrechte, Überwachung und Rückfallweg vor dem betrieblichen Einsatz klären.

DER FIKTIVE FALL: 95 % SENSITIVITÄT IM ERSTEN PoC

Bei Steinbach Metalltechnik erkennt ein extern getestetes Modell 95 % der nichtkonformen Teile im begrenzten Testdatensatz. Ob diese Leistung für andere Produktionsbedingungen trägt, bleibt offen. Die Kennzahl sagt zudem nichts über Fehlalarme oder realen Zeitgewinn.

DIE NÄCHSTE PRÜFUNG PRIORISIEREN

Im Steinbach-Fall zuerst B: unabhängige Evaluation am vorgesehenen Einsatz, schwierige Varianten gesondert prüfen. Nicht sofort weiter trainieren oder Schwellenwerte ändern. Ein möglicher wirtschaftlicher oder betrieblicher Ausschlussgrund kann diese Reihenfolge ändern.

Quellenbasis im Fachbeitrag: NIST SP 1270 · Suresh & Guttag (2021) · Nickerson (1998) · NIST AI RMF 1.0

Fünf Fehlerbereiche als redaktionelle Orientierung, keine normierte und trennscharfe Bias-Taxonomie.