

Information » KI Black Box «

Eine Black Box in der Künstlichen Intelligenz bezeichnet ein KI-System, dessen innere Abläufe und Entscheidungsprozesse für den Benutzer nicht nachvollziehbar oder erklärbar sind. Der Benutzer sieht lediglich die Ein- und Ausgaben des Systems, kann aber nicht verstehen, wie und warum die KI zu bestimmten Ergebnissen oder Entscheidungen kommt. Black Boxes in der KI werden mit einiger Skepsis betrachtet, da sie bedeutende Probleme mit sich bringen:

Der Mangel an Transparenz kann zu Misstrauen gegenüber KI-Systemen führen. Besonders in Bereichen wie Gesundheitswesen oder Bankwesen, wo Vertrauen in die Systeme essenziell ist.

In Szenarien, in denen KI wichtige Entscheidungen trifft oder Auswirkungen auf das Leben von Menschen hat, ist es schwierig, Verantwortlichkeiten zuzuordnen oder Fehler zu erkennen.

KI-Modelle können auch unbeabsichtigt Vorurteile aus Trainingsdaten übernehmen, wie das Beispiel des Amazon-Rekrutierungsalgorithmus zeigt, der Frauen systematisch benachteiligte.