

Daniel Feuerstack, Menschenrechtliche Vorgaben in Bezug auf die Intransparenz KI-basierter Entscheidungen, Beginn Juli 2022, geplanter Abschluss 2025

KI-Systeme und KI-basierte Entscheidungen sind oftmals intransparente „Black Boxes“. Das bedeutet, dass nicht oder nur mit großen Schwierigkeiten festgestellt werden kann, wie und aus welchen Gründen eine konkrete KI-basierte Entscheidung zustande kam. Diese Intransparenz, manchmal auch als fehlende Erklärbarkeit, Verstehbarkeit oder Nachvollziehbarkeit bezeichnet, resultiert zum einen daraus, dass manche KI-Systeme auf komplexen Modellen wie etwa neuronalen Netzen basieren, deren genaue Entscheidungsfindung für Menschen nicht nachvollziehbar ist. Zum anderen sind Informationen über die konkrete Entscheidungsfindung oftmals als Geschäftsgeheimnis geschützt. Dies ist vor allem dann problematisch, wenn KI-Systeme anhand personenbezogener Daten über Menschen entscheiden und betroffene Personen aufgrund dessen Fehlentscheidungen und Diskriminierungen nicht feststellen und vor Gericht beweisen können.

Nutzt der Staat ein KI-System, das auf komplexen neuronalen Netzen basiert oder dessen Entscheidungslogik als Geschäftsgeheimnis geheim gehalten wird, so kann dies gegen die grundrechtlich verankerte staatliche Begründungspflicht verstoßen. Diese ergibt sich insbesondere aus dem Grundrecht auf effektiven Rechtsschutz (Art. 19 Abs. 4 GG sowie Art. 2 Abs. 1 GG in Verbindung mit dem Rechtsstaatsprinzip). Auch Private in einer besonderen Machtstellung, etwa Kreditbewertungsdienste wie die SCHUFA oder Betreiber sozialer Medien, die KI-Systeme zur Bewertung von Personen nutzen, müssen nachteilige Entscheidungen diesen gegenüber begründen. Einfachrechtlich sind diese grund- und menschenrechtlich garantierten Transparenzanforderungen unter anderem in der Datenschutz-Grundverordnung und in der KI-Verordnung sowie in dem KI-Rahmenübereinkommen des Europarates verankert. Daniel Feuerstack untersucht und erörtert seit 2022 diese sich aus Menschenrechten und Unionsrecht ergebenden Vorgaben in seiner Dissertation und leitet daraus konkrete Handlungsvorschläge für die künftige KI-Forschung und -Praxis ab. Die Dissertation soll 2026 beendet werden.