



Prof. Dr. Anne Lauber-Rönsberg / Ass. iur. Franz Lehr

## Workshop: Transparenz von KI-Systemen

Termin: 10. Mai 2021

Technologien der Künstlichen Intelligenz (KI) werden bereits in vielen Bereichen von privaten Akteuren verwendet. Digitale Sprachassistenten oder Chatbots weisen ein Sprachverständnis auf, das selbstständige Kommunikation mit Kund:inn:en und Verbraucher:inne:n ermöglicht. Bei der Personalauswahl werden Bewerber:innen mittels KI-gestützten Scorings bewertet, ihre Daten und ihr anhand von Video- oder Tonaufnahmen analysiertes Kommunikationsverhalten werden einer Mustererkennung unterzogen, um der zu besetzenden Stelle bzw. den aus der Stellenbeschreibung entnommenen und für das System operationalisierten Anforderungen möglichst passgenau zu entsprechen. Individualisierte Produktempfehlungen und Zielgruppenanalysen zur Ansprache zu konvertierender potenzieller Kund:inn:en, KI-gestützte Credit-Scoring-Verfahren und Versicherungsabschlüsse, eine Unterstützung des Vertriebs von Verkaufspersonal durch digitale Assistenten, automatisiertes Aussteuern, Evaluieren und Anpassen von Marketingkampagnen – die (potenziellen) Einsatzfelder sind so vielfältig wie komplex.

Für die Betroffenen ist es häufig nicht ersichtlich, mit welcher Datengrundlage ein KI-System arbeitet, welche Operationen das System ausführt, wie Entscheidungen getroffen werden. Mitunter ist nicht einmal erkennbar, dass KI-Systeme zur Kommunikation eingesetzt werden. Besonders relevant wird dieser Umstand, wenn KI-Systeme eine Analyse persönlicher Dispositionen von Individuen und hierdurch eine personalisierte Anpassung der gewählten Kommunikation ermöglichen. Für eine solche Personalisierung der Kommunikation sind sicherlich viele sinnvolle Szenarien denkbar (z.B. die Identifizierung von Kommunikationssituationen, in denen Informationen in leichter Sprache übermittelt werden sollten). Möglich ist aber auch, dass diese Möglichkeiten gezielt zur Beeinflussung der Entscheidungsfindung von Individuen eingesetzt, wenn z.B. Individuen durch ein gezieltes Nudging empfänglicher für bestimmte Medien oder Vertragsangebote gestimmt werden.

In den letzten Jahren haben Staaten und Internationalen Organisationen, Unternehmen, NGOs und Forschungsinstitutionen eine Vielzahl an KI-Ethikleitlinien entwickelt. Fast alle verweisen dabei zentral auf den **Grundsatz der Transparenz**, die für die Vertrauenswürdigkeit von KI-Systemen grundlegend ist. Allerdings zeigt sich das Konzept der Transparenz ebenso wie damit zusammenhängende Begrifflichkeiten wie Erklärbarkeit, Nachvollziehbarkeit oder Rückverfolgbarkeit bislang wenig konturiert.

Zu beobachten ist zudem ein Fokus der Debatte auf den staatlichen Einsatz von KI-Systemen gegenüber den Bürger:inne:n. Im Verhältnis Staat - Bürger:inne:n, z.B. bei der Gesichtserkennung zur (Video-)Überwachung öffentlicher Räume, bei Predictive Policing oder bei KI-gestützter Leistungszuweisung im Sozialsystem, zielt die Forderung nach Transparenz insbesondere auf die Kontrolle demokratisch legitimierter Machtausübung.

Der uns primär interessierende privatwirtschaftliche Bereich ist dagegen dem Primat privatautonomer Lebensgestaltung gleichberechtigter Akteure unterworfen. Transparenz beruht hier notwendigerweise auf anderen Begründungsansätzen, die auch die Anforderungen an ihre Ausgestaltung verändern. Viel deutlicher wird hier die Funktion als Werkzeug für den Ausgleich von Informationsasymmetrien zur Ermöglichung von informierten, privatautonomem Entscheidungen der Vertragspartner:innen, der Zuordnung von Haftungsverantwortlichkeit und anderen Regulierungsansätzen. Aus privatrechtlicher Sicht wirft dies die Frage auf, ob der derzeitige Rechtsrahmen das ethisch begründete Transparenzgebot in hinreichendem Maße umsetzt.

Das Recht kann dieses Regelungsziel und seine Problemfelder nur adäquat adressieren, wenn es sowohl in ethischer und verhaltenswissenschaftlicher als auch in technischer Hinsicht auf fundierten Annahmen beruht.

**Im Rahmen des interdisziplinären Workshops sollen daher sowohl grundlegende als auch praktische Fragen nach der Bedeutung von Transparenz beim Einsatz von Künstlicher Intelligenz diskutiert werden: Warum, wie und durch bzw. für wen ist Transparenz zu gewährleisten? Welche Vor- und Nachteile bestehen bei unterschiedlichen Graden transparenter KI-Systeme? Ist Transparenz eine hinreichende oder gar notwendige Bedingung für Vertrauen der Betroffenen? Wieso wird in ähnlich gelagerten Fällen von KI-Systemen mehr Transparenz als von menschlichen Akteur:innen gefordert? Führen KI-Systeme zu der Verschleierung menschlicher Verantwortung? Wie kann Verantwortungsdiffusion vorgebeugt, wie kann Nach- bzw. Beweisbarkeit gesichert werden? Welche Kommunikationswege werden genutzt oder sind nutzbar, welche Rezeptionsgrade realistisch? Welchen Einfluss hat die Kommunikation mit Maschinen auf das menschliche Kommunikationsverhalten?**

Veranstaltet wird der Workshop im Rahmen des Teilprojekts „Transparenz als Voraussetzung für vertrauenswürdige KI“ des Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, das vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der KI-Strategie als Kompetenzzentrum für Künstliche Intelligenz gefördert wird.