Artificial (Un)Intelligence: Ausgewählte rechtliche Problemfelder

David Schneeberger, Pro Scientia Wien Vortrag 11. Jänner 2022

Künstliche Intelligenz (KI) darf nicht als etwas vom Menschen Losgelöstes, Artifizielles, Mysteriöses, das sich unserem Verständnis und somit auch rechtlicher Regulierung entzieht, missverstanden werden. KI-Systeme sind vielmehr Artefakte, die vom Menschen geschaffen und durch menschliche Grundsatzentscheidungen geprägt werden. Als Übergriff bezeichnet KI zahlreiche verschiedene Ansätze: Eine gegenwärtig häufig eingesetzte Technik ist das sogenannte Machine Learning (ML). Dabei wird zB für Spamerkennung aus großen Datenmengen (und häufig den „korrekten" Antworten wie Spam/kein Spam) mithilfe eines Algorithmus ein statistisches Modell erlernt (zB welche Wörter in Mails Spam konnotieren), das auf unbekannte Mails angewendet werden kann. Diese Modelle unterscheiden sich in vielen Aspekten zB welche Muster sie erkennen können, wie transparent sie sind oder welche „Stellschrauben" zur Verfügung stehen. Wichtig ist zu verstehen, dass ML-Systeme nicht nach menschlichen Mustern ein tieferes Verständnis „erlernen", sondern (häufig) statistische Oberflächlichkeiten aufgreifen. Teilweise bleibt dadurch unklar, ob ML-Systeme auch aus den richtigen Gründen zum richtigen Ergebnis kommen oder ob ein sogenannter „Kleverer-Hans-Effekt" vorliegt: ML neigt dazu hohe Performanz auf unsinnigen Wegen zu erreichen. So wurden Boote anhand von Wasser und Huskys anhand von Schnee erkannt ohne das eigentliche Objekt zu berücksichtigen.

Aber auch diese künstliche Dummheit stellt uns vor juristische Probleme: So unterminiert ML manche Grundannahmen des Datenschutzrechts. So können anonymisierte Daten, die nicht der DSGVO unterliegen, mittels ML zunehmend wieder Personen zugeordnet werden. So wurden die anonymisierten Bewertungsdaten von Netflix mit Blick auf die IMDB wieder den „Eigentümern" zugewiesen. Auch die Trennung zwischen „normalen" und „sensiblen" personenbezogenen Daten (zB Rasse, Geschlecht) löst sich zunehmend auf. So wurden in Experimenten Korrelationen zwischen Facebook-Likes und Intelligenz oder sexueller Orientierung gezogen. Betroffene haben somit kaum Einblick und Kontrolle über diese Verwendungszwecke. Ausgereifte Lösungen für diese Problematiken fehlen bislang. So wurden in der Informatik Verfahren wie k-anonymity oder differential privacy entwickelt, um Datensätze nutzbar zu machen und gleichzeitig das Risiko der Re-Identifizierung zu minimieren. Ebenso ist der Status von Schlussfolgerungen als personenbezogene Daten bislang strittig aber wohl zu verneinen und es wurde daher angeregt, ein Recht auf „vernünftige Schlussfolgerungen" zu schaffen.

Auch das Antidiskriminierungsrecht wird dabei vor neue Herausforderungen gestellt. So sind zB aufgrund ungleich repräsentativer Trainingsdaten kommerzielle Gesichtserkennungssysteme gerade für schwarze Frauen deutlich fehleranfälliger. In Österreich wurde der sogenannte AMS-Algorithmus dazu verwendet, um in Beratungsgesprächen die Vergabe von Betreuungsleistungen zu steuern. Er berücksichtigt weibliches Geschlecht, Betreuungspflichten und gesundheitliche Beeinträchtigung als negative Merkmale, die daher in weniger Förderung resultieren können. An den Grenzen experimentiert die EU bereits länger an der Befragung von Migranten mittels eines Avatars und ML „Lügendetektoren", die die Gefahr einer Täuschung zB über das Herkunftsland einstufen sollen. Rechtlich ergibt sich vor allem eine Beweislastproblematik: Ein Betroffener, der zB keine Beförderung erhält, muss beweisen, dass sich die Entscheidung auf ein rechtlich verbotenes Merkmal wie Geschlecht stützt. Das Recht hilft insofern, als es eine Beweislastumkehr vorsieht. Der Betroffene muss nur glaubhaft machen, dass ein Diskriminierungstatbestand vorliegt, danach muss der Arbeitgeber den Gegenbeweis antreten. In historischen Fällen war dies möglich. Aufgrund der sogenannten Black-Box-Eigenschaft von ML ist jedoch häufig unbekannt, ob solche Systeme verwendet werden, welche Merkmale sie einbeziehen, wie der Entscheidungsprozess abläuft. Ohne Zugangsrechte zu statistischen Daten und Fehlerraten gelingt dieser Beweis somit kaum. Gelingt er doch, dann hat der Beklagte als KI-Verwender Zugriff auf alle statistischen Daten und der Gegenbeweis dürfte leicht gelingen. Gleiches gilt auch für die Möglichkeit der Rechtfertigung; hohe Performanz könnte teilweise Diskriminierung rechtfertigen. Betroffene dürften sich somit rechtlich häufig in einer Sackgasse befinden. Problematisch ist auch die Existenz von Proxy-Diskriminierung. Ein Proxy ist eine Ersatzvariable für ein schwierig zu messende Variable. Bei ML kann jedoch eine scheinbar harmlose Variable ein rechtlich verbotenes Merkmal „kodieren": Stützt sich zB der Arbeitgeber auf Musikgeschmack als Merkmal müsste der Betroffene beweisen, dass dieses harmlose Merkmal ein Proxy zB für Geschlecht ist. Ohne aufwendige statistische Analysen dürfte das kaum gelingen. Umgekehrt droht die Gefahr, dass sich neue, vom Recht nicht geschützte Gruppen, herausbilden zB Katzenbesitzer oder Computer-Gamer. Ethisch mag es verwerflich sein nach solchen Gruppen zu differenzieren, rechtlich könnte man jedoch nicht von Diskriminierung sprechen. Ein vielversprechender Ansatz ist der von der Informatik getragene Bereiche Fairness in Machine Learning. Hier wird versucht externe, primär rechtliche, Grenzen, kodifiziert durch mathematische Fairnessdefinitionen, bei der Modellierung zu berücksichtigen, um Diskriminierung „herausfiltern" zu können. Jedoch ist bislang unklar, ob das in der Judikatur vertretene, nicht immer einheitliche und präzise Verständnis von Fairness, „automatisiert"

werden kann. Mathematisch bewiesen sind viele der möglichen Fairnessdefinitionen auch gegenseitig widersprüchlich.

Das Verständnis für ML-Systeme und damit die rechtliche Regulierung wird auch durch die Black-Box-Eigenschaft erschwert. Diese Metapher beschreibt den Umstand, dass Input und Output beobachtet werden können, aber unklar bleibt, wie der Entscheidungsprozess abläuft. Intransparenz ist dabei keine intrinsische Eigenschaft von ML. Manche komplexen Modelle wie Künstliche Neuronale Netze setzen sich jedoch aus tausenden simpleren mathematischen Einheiten und Verbindungen zusammen die nur im Zusammenspiel funktionieren. Abgesehen von dieser technischen Seite unterliegen viele Systeme, die von Unternehmen entwickelt werden, Geheimnisschutz wie Geschäfts- oder Betriebsgeheimnissen. Diese Problematik hat in vielen Bereichen zu intensiven rechtlichen Debatten geführt. So wird Datenschutzrecht seit Jahren diskutiert, ob ein „Recht auf Erläuterung“ existiert. Auch nach Jahren ist die Existenz strittig. Im Produktsicherheitsrecht zB bei Medizinprodukten ist es fraglich, ob man von Sicherheit sprechen kann, ohne genau darüber Bescheid zu wissen, wie ein solches System funktioniert oder wann es nicht mehr funktionieren wird. Ähnliche Fragen stellen sich im ärztlichen Berufsrecht zB bei der Aufklärung von Patienten. Kann ein Patient entsprechend aufklärt werden, wenn der Arzt selbst ein ML-System nicht versteht? Gerade beim staatlichen Einsatz stellt sich die Frage, ob das rechtsstaatliche Prinzip die Black-Box-Eigenschaft begrenzt. So verwendet Österreich ein ML-System zur Analyse von Steuererklärungen, um das Risiko der Täuschung zu eruieren und Nachprüfungen zu koordinieren. Die genauen Mechanismen werden bewusst verborgen, um Steuerhinterziehung zu verhindern. Dadurch ist unbekannt, welche Informationen einbezogenen werden und ob dieses System auch regelmäßig überprüft wird. Um diese Black Box zu öffnen, wird in der Informatik schon länger an sogenannter explainable AI (xAI) geforscht.

Die bisher dargestellten Rechtsbereiche wurden nicht mit Blick auf KI konzipiert. Um etwaige Lücken zu schließen präsentierte die Europäische Kommission im April 2021 den Vorschlag für eine europaweit einheitliche Regulierung, den Artificial Intelligence Act (AIA). AIA basiert auf einer Risikopyramide, Systeme auf der höchsten Risikostufe wie biometrische Überwachung, social scoring nach dem Vorbild Chinas sind verboten. Hochrisikosysteme unterliegen einer strikten Regulierung. Wie auch im normalen Produktsicherheitsrecht ist geplant, dass KI-Systeme einem Konformitätsbewertungsverfahren unterliegen bevor sie auf den Markt gebracht werden können.

Quellenverzeichnis:

Alpaydin, Machine learning. The new AI (2016).

Barocas/Selbst, Big Data's Disparate Impact, California Law Review 2016, 671–732.

Kearns/Roth, The Ethical Algorithm. The Science of Socially Aware Algorithm Design (2020).

Lapuschkin et al, Unmasking Clever Hans predictors and assessing what machines really learn, Nature communications 2019, 1096.

Marcus/Davis, Rebooting AI. Building Artificial Intelligence We Can Trust (2019).

Molnar, Interpretable Machine Learning. A Guide for Making Black Box Models Explainable[2] (2022), https://christophm.github.io/interpretable-ml-book/index.html#.

Orwat, Diskriminierungsrisiken durch Verwendung von Algorithmen. Eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes (2019).

Russell/Norvig, Artificial Intelligence. A Modern Approach[4] (2021).

Veale/Zuiderveen Borgesius, Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach, CRi 2021, 97–112.

Wachter/Mittelstadt, A Right to Reasonable Inferences, Columbia Business Law Review 2019, 494–620.

Zarsky, Incompatible. The GDPR in the Age of Big Data, Seton Hall Law Review 2017, 995–1020.