# Wenn die KI “A” sagt und die Ärztin “B”

## Verantwortungsverteilung im Umgang mit algorithmischen Helfern

**Roman Lehner**

24.11.2022 doi: 10.17176/20221124-121631-0

Wenn die medizinische Diagnose-KI „A“ vorschlägt, sagt die Ärztin dann noch „B“? Oder anders gewendet: Kann sie vernünftigerweise noch „B“ sagen, wenn sie im Fall, dass sich „A“ als richtig herausstellt, für die Fehldiagnose „B“ haften müsste? Wenn hingegen „A“ sich als Fehldiagnose entpuppt und die Ärztin auch „A“ gesagt hat, ist die Lage anders, denn dann haftet – womöglich – auch die KI. Und wenn die Ärztin „A“ sagt, nachdem die KI „B“ vorgeschlagen hat und „A“ auch richtig war, bekommt sie vermutlich eine Auszeichnung. Ob wir letzteres jemals erfahren werden, hängt davon ab, ob es diagnostische Abweichungen von der KI-Empfehlung in der Praxis überhaupt noch geben wird. Und, Hand aufs Herz, wie würden Sie es handhaben: Erst die eigene Diagnose oder erst den KI-Vorschlag? Der Gefahr der narzisstischen Selbstkränkung entgeht man am besten, wenn man erst die KI befragt und dann deren Empfehlung mit den Worten übernehmen kann: So hätte ich es auch gesehen!

### Herausforderungen, Chancen, Potenziale

Zentral für die Frage nach einem adäquaten rechtlichen Umgang mit KI ist die Rollenverteilung zwischen Mensch und technischem System. Wer ist hier Herr, wer Knecht? Wer entscheidet in letzter Konsequenz und muss dann aber auch für die Folgen der Entscheidung geradestehen? Medizinische Diagnose-Assistenz-Systeme sollen Therapievorschläge gestützt auf diagnostische Wahrscheinlichkeitsberechnungen unterbreiten, sie sollen, wie es im Namen auch angelegt ist, assistieren und nicht den menschlichen Letztentscheider zum Gehilfen degenerieren lassen. Susanne Beck hat kürzlich den im einschlägigen Diskurs prominenten Begriff des „Haftungsknechts“ für diese Situation ins Spiel gebracht, als sie auf einer Konferenz zum Thema „Künstliche Intelligenz – Herausforderungen und Chancen“, veranstaltet von der Deutschen Sektion der Internationalen Juristen-Kommission in Göttingen, über die rechtlichen Probleme des KI-Einsatzes im Gesundheitsbereich referierte. Dass die Tagung, die sich im Kern um die Frage nach Regulierungsbedarf und Regulierungsmöglichkeiten von KI in verschiedenen Lebensbereichen widmete und eine erhebliche Resonanz erfahren hat (hier und hier), mit den „Herausforderungen“ und „Chancen“ des KI-Einsatzes den Spannungsbogen gut getroffen hat, zeigt auch der aktuelle (informelle) Trilog zur KI-Verordnung der EU. Der im April dieses Jahres vorgelegte Verordnungsentwurf betont zum einen, dass KI-Techniken „zu vielfältigem Nutzen für Wirtschaft und Gesellschaft über das gesamte Spektrum industrieller und gesellschaftlicher Aktivitäten hinweg beitragen können“ (EG Nr. 3). Zum anderen werden auch die technologischen Risiken hervorgehoben (EG Nr. 4), die eine rechtliche Regulierung dann eben auch besonders nahelegen. Chancen und Herausforderungen eben, wobei die Chancen immerhin als erstes genannt werden – das muss auf einer juristischen Tagung erwartungsgemäß umgekehrt sein.

Tatsächlich geht die Kommission davon aus, dass KI in fast allen Lebensbereichen von enormem Nutzen sein kann, ihr Einsatz kann „zu guten Ergebnissen für Gesellschaft und Umwelt führen, beispielsweise in den Bereichen Gesundheitsversorgung, Landwirtschaft, allgemeine und berufliche Bildung, Infrastrukturmanagement, Energie, Verkehr und Logistik, öffentliche Dienstleistungen, Sicherheit, Justiz, Ressourcen- und Energieeffizienz sowie Klimaschutz und Anpassung an den Klimawandel“ (EG Nr. 3). Ebenso groß erscheinen aber auch die Risiken, zumindest die Risikowahrnehmungen. Bei dem unter der tschechischen Ratspräsi-

dentschaft erarbeiteten Kompromissvorschlag war gerade die Tinte getrocknet, da beeilte sich manch einer schon, dystopische Zustände abwehren zu müssen. Von einer geplanten „biometrische[n] Massenüberwachung im öffentlichen Raum“ ist nun plötzlich die Rede, dabei wurde gegenüber dem Kommissionsentwurf dem Erwägungsgrund Nr. 19 lediglich eine Klarstellung dahingehend beigefügt, dass es Sicherheitsbehörden prinzipiell möglich sein soll, KI-gestützte Echtzeitidentifikationssysteme auch nach erfolgter EU-Harmonisierung einzusetzen. In Bezug auf die Vorratsdatenspeicherung hat sich der EuGH (inzwischen) darauf eingelassen, dieses technische Verfahren zumindest in eng umgrenzten Ausnahmefällen zuzulassen (hier, hier und hier), was unterstreicht, dass Totalächtungen in die falsche Richtung weisen. Dass KI-Regulierung nicht KI-Verunmöglichung sein und dass überhaupt Überregulierung vermieden werden sollte, hat kürzlich der Branchenverband der Digitalwirtschaft offensiv eingefordert und vor einem (weiteren) Zurückfallen Europas und insbesondere Deutschlands im internationalen Wettbewerb gewarnt. Böse Zungen behaupten, die EU habe sich nur deshalb das Ziel gesteckt, „bei der Entwicklung einer sicheren, vertrauenswürdigen und ethisch vertretbaren künstlichen Intelligenz weltweit eine Führungsrolle einzunehmen“ (EG Nr. 5), weil im technologischen Feld selbst eine solche Rolle ernsthaft nicht mehr angestrebt werden könne. Insofern sind Debatten um Totalächtungen bestimmter KI-Systeme, wie gerade im Bereich der Gesichtserkennung im öffentlichen Raum zu beobachten, auch ein Indiz dafür, dass in Europa geradezu notorisch über KI-Begrenzung und viel weniger über KI-Potenziale nachgedacht wird. Ähnlich gelagerte Debattenlagen wie die zum Einsatz KI-gestützter tödlicher Drohnen zeigen indes, dass die einschlägigen normativen und ethischen Fragen im Kern vor allem auf einen Punkt zulaufen: Dürfen wir zulassen, dass am Ende eine Maschine *anstelle* eines Menschen Entscheidungen (in letzter Konsequenz sogar über Leben und Tod) trifft?

## Autonome Systeme, autonome Entscheidungen?

Beim autonomen Fahren wie bei anderen autonom agierenden Systemen soll der „human in the loop“ zumindest in Problemsituationen als Letzthandelnder zum Einsatz kommen. Im Zweifel wird der Kraftfahrer, der von seinem autonomen Auto plötzlich das Signal bekommt, dass dieses überfordert ist und die Kontrolle an den Menschen zurückzugeben wünscht, nicht nur weniger aufmerksam sein als er wäre, wenn er von Anfang an das Fahrzeug selbstständig gelenkt hätte. Das ganze Modell autonomer Systeme steht in Frage, wenn die menschliche Letztkontrolle, zumindest in kritischen Momenten, als normativer Rettungsanker bereitstehen soll. Wie Gerald Spindler auf der Göttinger Tagung in Bezug auf Modelle der Vernetzung selbstfahrender Fahrzeuge mit externen Schaltzentralen, in denen dann Menschen sitzen sollen, geradezu süffisant anmerkte, könne man von autonomem Fahren eigentlich nur sprechen, wenn infolge eines Funklochs die Verbindung zur Zentrale abbreche. Denn dann wäre das Fahrzeug, vom Insassen einmal abgesehen, tatsächlich auf sich allein gestellt. Das Desiderat des Informatikers Simon Burton ging dann auch in eine andere Richtung: Die Sicherheit beim Einsatz von KI könne ohnehin nur schrittweise erhöht werden und dementsprechend könne auch die Regulierung sinnvoll nur in Etappen erfolgen.

Daran, selbstlernenden Systemen, wenn man sie denn einsetzen will, im Grundsatz auch zu ‚vertrauen‘, kommt man aus technischer Sicht offenbar nicht vorbei. Das zeigt im Übrigen auch das bekannte „black-box“-Problem, das die Rekonstruktion einer KI-‚Entscheidung‘ im Nachhinein und somit auch die individuelle Bestreitbarkeit in aller Regel ausschließt oder zumindest ganz massiv erschwert. Der EuGH hat in seiner Entscheidung zur Fluggastdatenspeicherung genau aus diesem Grund KI-autonomen Verwaltungseingriffen einen Riegel vorgeschoben. Auch hier wird offenkundig davon ausgegangen, dass in letzter Konsequenz nur ein menschlicher Entscheidungsträger akzeptabel sein soll, vor allem deswegen, weil dieser, anders als die KI, seine Entscheidung rechtsschutzermöglichend begründen kann. Daraus folgt, dass

Menschen beim Einsatz von KI in grundrechtssensiblen Fragen am Ende die letzte Kontrolle auszuüben haben, was vor allem prozedural sichergestellt werden muss. Eine pauschale Ächtungsattitüde hilft auch hier nicht weiter. Zu wenig wird etwa darüber nachgedacht, dass KI-gestützte Eingriffe gegenüber rein menschlichen Eingriffen im Einzelfall ein milderes Mittel darstellen könnten, wie Eric Hilgendorf in Göttingen hervorhob und am Beispiel eines (automatisierten) Flughafen-Nacktscanners verdeutlichte. Zu sehr wird überdies die menschliche Entscheidungsrationalität idealisiert. Dass der menschliche Entscheidungsprozess, der dem Verfassen einer Begründung vorgelagert ist, selbst eine „black box" darstellt, wird gemeinhin wie selbstverständlich hingenommen. Hier zeigt sich etwas, was die Debatte um KI generell prägt: Allgemein akzeptabel scheint der Einsatz algorithmisierter Entscheidungsverfahren offenbar nur dann zu sein, wenn die KI (deutlich) weniger Fehler macht als der Mensch.

## „Doctor in the loop" als „Haftungsknecht"?

Umgekehrt ist aber auch eine evident menschenüberlegene KI problematisch. Hier geht es um das Problem, dass menschliche Entscheidungen durch (besonders gute) KI-Entscheidungsempfehlungen faktisch determiniert werden. Wer sich gegen die KI entscheidet, wird dies als diagnostizierende Ärztin eben besonders, vielleicht sogar übermäßig gut begründen müssen. Zu den Risiken in diesem Zusammenhang gehört auch das sogenannte „automation bias", also die mögliche Perpetuierung struktureller Ungleichheiten etwa im Bereich der Berechnung von Rückfallwahrscheinlichkeiten für Strafaussetzungsentscheidungen oder bei der Vorsortierung von Bewerbungsunterlagen in Stellenbesetzungsverfahren. Denn die KI schaut nur zurück, sodass etwa eine diskriminierende Einstellungspraxis in der Vergangenheit von der KI nicht nur aufrechterhalten, sondern sogar optimiert werden könnte. All diese Vorprägungsrisiken sind es, derentwegen Martin Eifert in Göttingen die Forderung nach einem generellen Gesetzesvorbehalt für den Einsatz von KI in relevanten Entscheidungssituationen zur Diskussion stellte, also unabhängig vom Vorliegen eines menschlichen Letztentscheiders.

Den geplanten Ansatz der EU, zumindest bei Hochrisiko-KI-Systemen auf erhebliche Beweiserleichterungen zu setzen, wirkt vor dem Hintergrund der faktischen Determination menschlicher Entscheidungen allerdings auch ein wenig defätistisch. Denn wenn der Mensch als Nutzer der KI kein reiner „Haftungsknecht" sein soll, bleibt nur der Rückgriff auf den Inverkehrbringer des Systems. Die Haftungsschäden für Fehldiagnosen werden dann aber von den KI-Unternehmen wie Entwicklungskosten schlicht eingepreist werden, jedenfalls zivilrechtlich. Denn strafrechtlich haftet in Fällen, in denen die Ärztin dem hochintelligenten Diagnose-Assistenzsystem folgt, dies zum Behandlungsfehler führt und der Patient stirbt, in aller Regel niemand. Der „doctor in the loop" hat wohl kaum sorgfaltswidrig gehandelt, wenn er einem System vertraut hat, das nachweislich radiologische Aufnahmen besser auswerten kann als er selbst, und den auf einen individuellen Programmierer zurückführbaren Programmierfehler, der nachweislich kausal und zurechenbar zum Tod des Patienten geführt hat, wird es vermutlich nicht geben.

## Die KI ist (noch) kein Mängelwesen

Die KI schlägt vor, der Mensch ‚entscheidet' und verantwortlich ist am Ende niemand? Die Annahme jedenfalls, dass, wenn die KI „A" sagt, der KI-Nutzer als menschlicher Letztentscheider im Einzelfall auch einmal für „B" votiert, ist tendenziell unrealistisch und die KI entscheidet dann faktisch eben doch ‚autonom', ohne dass allerdings dahinter ein autonomes Subjekt stehen würde. Die Philosophin und Publizistin Thea Dorn, die auf der Göttinger Tagung ebenfalls einen Auftritt hatte, hat jedenfalls zu Recht starke Zweifel daran angemeldet, dass der Versuch des Menschen, durch den Einsatz von KI seine eigene Mängelhaftigkeit, die sich ja auch in den Mängeln menschlichen Handelns und Entscheidens zeigt, real zu überwinden, erfolgreich sein wird. Ganz offenkundig ist die Mängelhaftigkeit des Individuums die Kehrseite (vielleicht gar die Voraussetzung?) von Verantwortung und Autonomie als Subjekt. Thea Dorn hat es auf den Punkt

gebracht: Von Autonomie könne im Kontext von KI erst die Rede sein, wenn der Pflegeroboter eines Tages beschließt, fortan lieber als selbstfahrendes Auto arbeiten zu wollen. Bis dahin jedenfalls werden wir an der menschlichen Verantwortung für all die politisch in den Blick genommenen Bereiche – „Gesundheitsversorgung, Landwirtschaft, allgemeine und berufliche Bildung, Infrastrukturmanagement, Energie, Verkehr und Logistik, öffentliche Dienstleistungen, Sicherheit, Justiz, Ressourcen- und Energieeffizienz sowie Klimaschutz und Anpassung an den Klimawandel“ – nicht vorbeikommen.

AUTOR*IN
**Roman Lehner**
Dr. Roman Lehner ist Privatdozent an der Georg-August-Universität zu Göttingen.

EMPFOHLENE ZITATION
Lehner, Roman: Wenn die KI “A” sagt und die Ärztin “B”: Verantwortungsverteilung im Umgang mit algorithmischen Helfern, Verfassungsblog, 24.11.2022, doi: 10.17176/20221124-121631-0.

Lizensiert unter CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20221124-121631-0.