



Künstliche Intelligenz in der Krebsforschung und Biomedizin - Artificial Intelligence in Cancer Research and Biomedicine

Authors: Jona Boeddinghaus
Submitted: 20. February 2021
Published: 1. March 2021
Volume: 8
Issue: 1
Affiliation: PSIORI
Languages: German
Keywords: Artificial Intelligence, Cancer, Research, Biomedical
Categories: Artificial Intelligence, Modeling and Simulation, Medicine
DOI: 10.17160/josha.8.1.739

Abstract:

Artificial Intelligence opens up huge possibilities in a variety of sectors. Imagine that doctor reports can be evaluated mechanically, complex, even unstructured data of hospital information technology can be derived in content specific information. In regard to the AI approval and publication, an ethical approach is of utmost importance for further development. If we follow the principles of trustworthy ethical AI, then AI will make a decisive contribution to making the lives of many people better. Presentation has been held on October 22nd, 2020 at the Symposium "KI and Krebs - Erkenntnisgewinnung in der Krebsforschung durch künstliche Intelligenz" at Haus der Industrie in Vienna. AF Institute: <https://www.af-institute.at>

JOSHA

[josha.org](https://www.josha.org)

**Journal of Science,
Humanities and Arts**

JOSHA is a service that helps scholars, researchers, and students discover, use, and build upon a wide range of content



Jona Boeddinghaus

.....
Künstliche Intelligenz
in der **Krebsforschung**
und **Biomedizin**
.....

AF INSTITUTE
Working Paper 2020/07



www.af-institute.at

Es ist offensichtlich, dass die KI-Forschung und -Anwendung in der jüngsten Zeit enorme Fortschritte gemacht hat. In KI-Methoden und Algorithmen liegt ein riesiges Potenzial.

Die Disziplin KI ist in einer Phase, in der sie erwachsen wird, würde ich sagen. Die Forschung und die Algorithmen gibt es seit Mitte des vorigen Jahrhunderts – das wird oft übersehen. Schon damals gab es viele interessante Möglichkeiten, Expertensysteme und Regelsysteme, die alle auch KI waren.

Maschinelles Lernen

Was jetzt neu ist und die KI beflügelt, ist eine Teildisziplin der KI: Maschinelles Lernen (ML). ML geht von Daten aus und lernt aus Daten. Es gibt drei Gründe, warum ML jetzt so gut funktioniert, zwei davon sind offensichtlich:

- Zum einen haben wir heute großen Datenmengen.
- Hinzu kommen leistungsstarke technische Systeme und standardisierte Datensammlungen, wie sie heute verbreitet sind (*data warehouses*). Das gab es vorher nicht. Erst die leistungsstarken Rechner von heute sind imstande, diese Datenmengen zu verarbeiten. Das ist bei KI besonders wichtig, denn beim Maschinellen Lernen müssen diese Daten im Zuge des Trainingsprozesses immer wieder und wieder verarbeitet werden. Es braucht also außerordentlich starke und schnelle Hardware.
- Der dritte Punkt – und das wird oft übersehen – sind die enormen Fortschritte, die die Computerwissenschaft seit den 1950er-Jahren errungen hat. Das betrifft nicht nur die Daten und die *compute power*, sondern es sind tatsächlich auch die Computerwissenschaftler und Informatiker, die trotz vieler Rückschläge Methoden entwickelt haben, die es heute erlauben, mit solchen großen Datenmengen auch umzugehen.

Das sind, ohne zu sehr ins Detail zu gehen, insbesondere die tiefen neuronalen Netze. Immer öfter wird über diese berichtet. Sie sind einfach eine bestimmte, besondere Form des Maschinellen Lernens, die mit sehr viel

größeren Datenmengen umgehen kann. Hinzu kommt, dass neuronale Netze viel robuster mit Daten umgehen, d.h. sie lassen sich nicht so leicht aus der Bahn werfen, wenn Daten nicht ganz gleichförmig aufbereitet sind.

Gelungene Beispiele

Das AF Institut kann gelungene Beispiele dazu vorweisen, in der Anwendung von KI für die Krebsforschung, aber auch für die Praxis. Nachstehend eine kleine Auswahl:

In **Abb. 1** sehen wir links oben ein Schaubild des Krebsregisters, einer Krebsdatenbank, die im AF-Institut erstellt wurde – aus dem Bedürfnis, einfach überhaupt eine Datensammlung zu standardisieren.

Nur wenn wir strukturiert Daten aus der Onkologie erfassen, können wir sie mathematischen Algorithmen überantworten und Informationen extrahieren. Diese Krebsdatenbank liefert also standardisierte Schnittstellen zu anderen Systemen und bietet die Möglichkeit, die gesammelten Daten auch für die Analyse zu nutzen.

Abb. 1 zeigt links unten einen „bunten Kugelhaufen“. Das ist ein sogenannter Cluster, eine Visualisierung eingebetteter Daten. Genexpressionen von mehreren Tausend Patienten wurden von einem tiefen Netz gelernt und eingebettet. Dieses neuronale Netz hat quasi die räumliche Ähnlichkeit der Datenpunkte gelernt – nur anhand der Genexpression! Im Anschluss wurden die zusammengehörigen Punkte eingefärbt.

Die Farben codieren die Diagnosen, die Krebstypen. Was die Grafik gut zeigt: Die räumliche Anordnung dieser Punkte fällt ziemlich gut mit der Farbcodierung zusammen. Das bedeutet: Das Netz hat selbstständig – nur aus den vorhandenen Daten – gelernt, diese Daten so zu gruppieren, dass sie mit der Diagnostik übereinstimmen.

Im Umkehrschluss heißt dies, dass man Ausreißer – also Falschdiagnosen oder auch Untergruppierungen – erkennen könnte. Onkologen und andere medizinische Experten können das Ergebnis dann bewerten und Entscheidungen treffen – z.B., ob weiter untersucht und geforscht werden soll oder ob die Diagnostik nochmal überdacht werden muss.

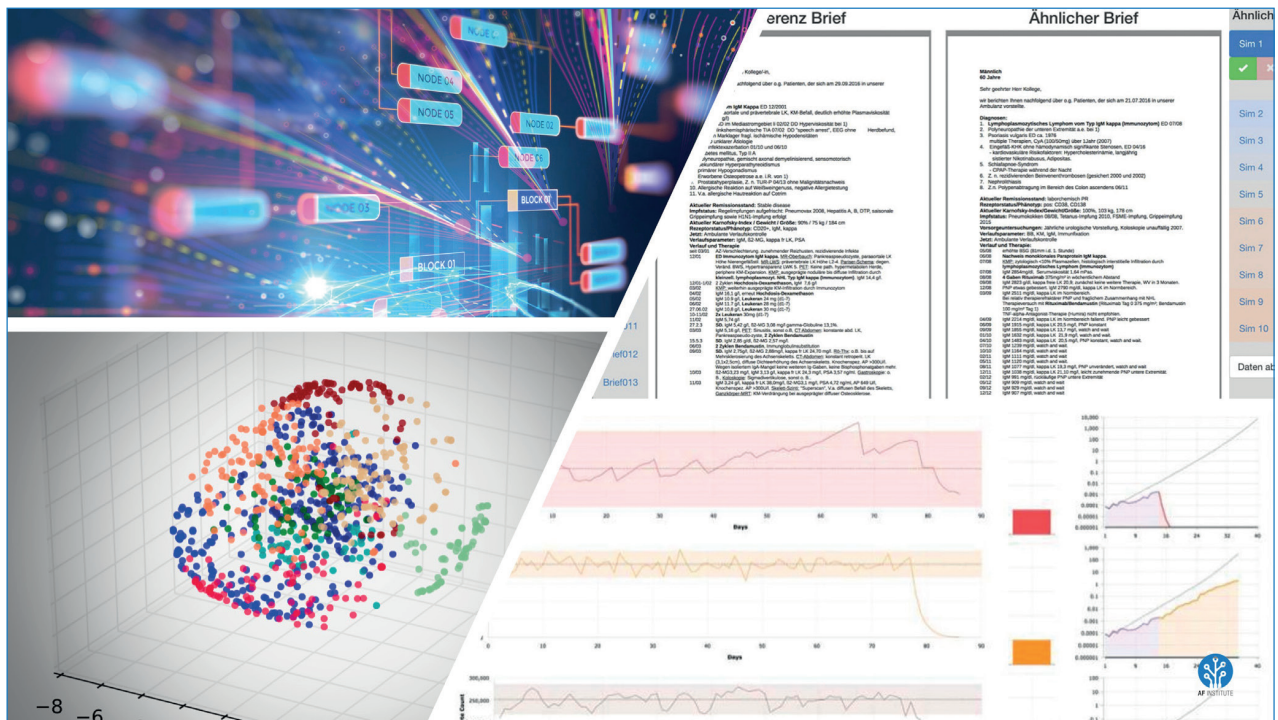


Abbildung 1

Weiters zeigt **Abb. 1** rechts unten ein Bild eines Blutmodells. Genauer: Es ist ein Hämatopoiesemodell, eine mathematische Formulierung des Blutwachstums. Wir gehen davon aus, dass wir mit mathematischen Modellen mehr über die Prozesse des Blutwachstums erfahren. Versuche im Labor sind Handarbeit, sie kosten viel Zeit. Die mathematische Formulierung medizinischer Probleme führt zu einer drastischen Verkürzung des Zeitaufwands.

Oben rechts sieht man in **Abb. 1**, wie von uns entwickelte KI Ähnlichkeiten in Dokumenten erkennt. Wir haben ein Verfahren zur maschinellen Auswertung von Arztbriefen entwickelt. Das ist ein ganz spannender Punkt, der auch schon mit vergleichsweise wenigen Daten gut funktioniert. Ja, Maschinelles Lernen braucht im Grunde immer Big Data; aber, wie dieses Beispiel zeigt, es gibt auch Anwendungen, bei denen Künstliche Intelligenz schon bei verhältnismäßig wenigen Ausgangsdaten gut funktioniert.

Bei den Arztbriefen hat unsere Software anhand weniger Daten Sprachmodelle gelernt. Diese Sprachmodelle erlauben uns bzw. dem Anwender, einen Arztbrief – in unserem Fall Briefe der Onkologie – der Software zu

übergeben. Die Software präsentiert dann bis zu zehn ähnliche Arztbriefe.

Worin liegt der Nutzen?

Was ist der Nutzen solcher Entwicklungen? Wir haben diese Algorithmen entwickelt, um Ärzte in ihrer medizinischen Arbeit zu unterstützen.

Viele Therapien schöpfen aus Erfahrung. Es kann durchaus interessant sein zu sehen, wie andere Mediziner bei ähnlichen Fällen vorgegangen sind, welche Parameter und Verläufe es bei Vergleichsfällen gab.

Wenn KI doch so toll und erfolgreich ist und unendlich viele Möglichkeiten bietet, stellt sich eine wichtige Frage: Warum ist KI noch längst nicht überall ausgerollt und im produktiven Einsatz?

Gerade so kritische und wichtige Bereiche wie die Onkologie erfordern eine behutsame Vorgehensweise, vor allem die richtige Fragestellung.

KI wird oft als magische Blackbox verstanden: Man gießt ein paar Daten hinein, schüttelt dann kräftig und am Ende kommt etwas ganz Tolles dabei heraus. So ist es aber nicht. KI kann nur dann funktionieren, wenn die Frage, die die KI-Lösung beantworten soll,

richtig gestellt wird. Und es muss ganz klar wissenschaftlichen Prozessen gefolgt werden, um diese Methoden zu implementieren.

Junge Disziplin

Ein weiterer Grund, warum es KI in der Praxis momentan noch besonders schwer hat: KI ist eine relativ junge Disziplin, sie stößt noch auf viele Vorbehalte.

Skeptiker vermuten, dass Ärzte durch Maschinen ersetzt werden und Jobs verloren gehen.

Es wird befürchtet, dass Maschinen in Prozesse eingreifen und nachteilige Entscheidungen treffen.

Hinzu kommen Anforderungen an den Datenschutz, gerade im medizinischen Umfeld. An den Kliniken ist der Schutz von Patientendaten ein Riesenthema – völlig zu Recht. Andererseits: Wenn wir datengetriebenen Algorithmen aus Datenschutzgründen nicht auf Daten anwenden können, hilft uns Künstliche Intelligenz nicht weiter.

Manchmal sind aber auch die Erwartungen an Künstliche Intelligenz zu hoch. Künstliche Intelligenz kann nicht jede Automatisierung erbringen. Viele Ziele sind mit den aktuellen Methoden gar nicht erreichbar.

Aus diesen und anderen Gründen steckt diese junge Branche weitgehend noch in Pilotprojekten. Scherzhaft ließe sich sagen, dass

in der KI mehr „Piloten“ zu finden sind als bei großen Fluglinien.

Nach den Piloten

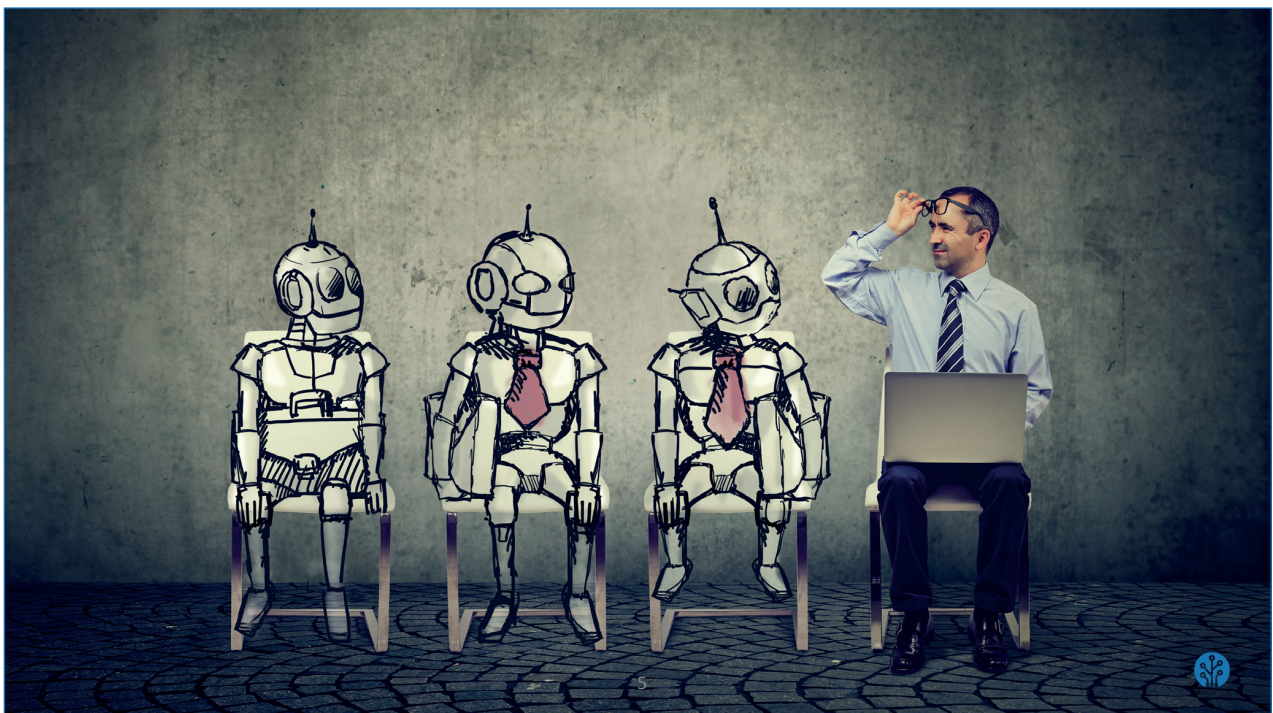
Das führt uns zur Situation, dass es im Bereich der KI viele spannende *Ideen*, sehr viele *Versuche* und eben *Pilotprojekte* gibt. Die große Herausforderung ist es, das alles auf Schiene zu bringen, zu operationalisieren und umzusetzen.

Wie lässt sich das lösen und vor welchen Problemen stehen wir dabei?

Zunächst gilt es grundsätzlich zu verstehen, dass *machine learning* als die wichtigste Disziplin der KI auf Daten basiert. ML ist nichts Anderes als Algorithmen, die aus Erfahrung lernen. Somit folgen diese Algorithmen nicht einprogrammierten Regeln, sondern sie lernen aus Erfahrung, und zwar in dem Sinne, dass sie von Trainingsdaten lernen. Auf Basis dieser Trainingsdaten – also Daten, die dem Algorithmus übergeben werden – trifft die Maschine Entscheidungen oder Vorhersagen.

KI braucht Regulierung

Das bedeutet aber auch: Wenn wir mit diesen Daten umgehen möchten und die Algorithmen vernünftig in Position stellen und tatsächlich anwenden wollen, dann müssen wir ein Regelwerk dafür definieren. Folglich sind es Menschen und nicht Maschinen, die die



Kriterien vorgeben, nach denen KI oder die Maschine lernt. Der Mensch muss also definieren, wie die Algorithmen funktionieren sollen und wonach sie sich richten sollen, damit erfolgreich und vertrauensvoll gearbeitet werden kann.

Wir werden also früher oder später Themen der Regulierung diskutieren müssen. Das mag zunächst als lästiges Hindernis wahrgenommen werden, aber diese Regulierung ist allein schon deshalb wichtig, weil mit ihr – quasi im Schlepptau – die Standardisierung kommt. In der Folge funktionieren Prozesse einfacher: Ich kann mich an Regeln halten und bin dann sicher, dass sie sauber durchgeführt werden können; ich kann sie überprüfen und ich kann sicher sein, dass die Dinge, die entwickelt wurden, tatsächlich auch das tun, was sie tun sollen.

Gerade in den Ingenieurwissenschaften und den kommunizierenden Gewerben ist es gar nicht anders denkbar.

Wir können nicht einfach irgendetwas machen, ohne Standardprozessen zu genügen. Die entwickelten Produkte oder Prozesse würden nicht funktionieren. Das ist ein ganz wichtiger Punkt. Erst wenn KI an diesem Punkt angekommen ist, wird sie wirklich erwachsen sein. Erst dann wird sie großflächig einen aktiven Beitrag leisten können. Das gilt auch für den Einsatz in der Onkologie: KI

muss Regeln folgen und sie muss der Regulierung näherkommen.

KI und Ethik

Im Folgenden möchte ich näher auf vertrauenswürdige KI eingehen. Ethik spielt dabei eine ganz wichtige Rolle.

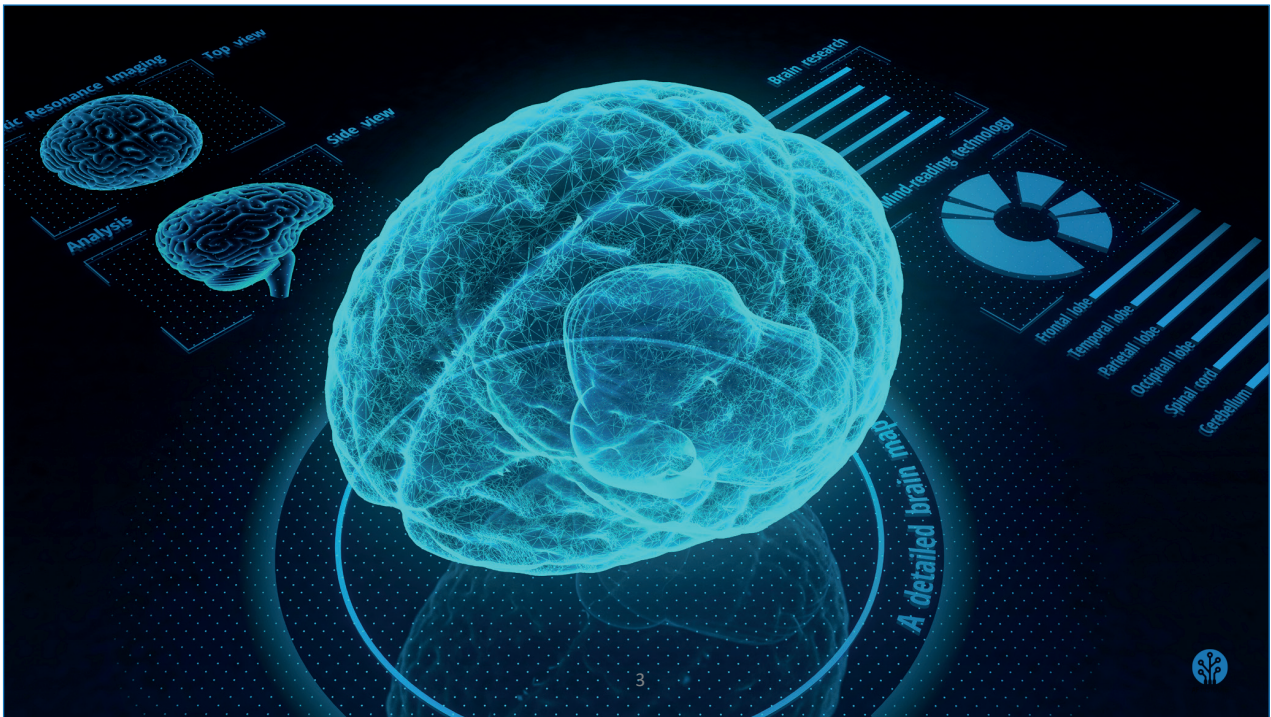
Ethik kann natürlich nur vom Menschen kommen. Ethik an sich ist ein philosophisches Prinzip: Was ist gut? Was ist richtig? Antworten darauf können nur Menschen geben, selbst „pflichtbewusste“ Maschinen wären dazu nicht imstande.

Stimmen gewisse Prognosen, dann wird KI irgendwann so intelligent – im Sinne von leistungsfähig – werden, kognitive Aufgaben zu lösen. Aber werden Maschinen so etwas wie ein ethisches Bewusstsein oder Moral entwickeln? Trotz aller Begeisterung für KI und Zuversicht hinsichtlich des technischen Fortschritts glaube ich das nicht.

Ethische Regeln müssen von den Menschen kommen. Ihre Aufgabe ist es, KI-Anwendung so zu gestalten, dass diese Anwendung eben genau diesen ethischen Regeln genügen.

Zu Fragen von KI und Ethik gibt es bemerkenswerte Vorschläge der OECD und der EU. Beide Organisationen kamen in Arbeitsgruppen zum Schluss, dass es eine Art Generalre-





gel geben muss: Der Mensch kommt zuerst; der Mensch muss immer die Aufsicht über die Maschine haben; er muss über die KI-Algorithmen das letzte Wort haben.

Transparenz

Wer sich mit Ethik und KI beschäftigt, wird rasch bei Fragen der Transparenz und Zurechenbarkeit ankommen. Alles, was KI macht, muss nachvollziehbar und der KI und dem Algorithmus zuschreibbar sein.

In seiner Konsequenz gar nicht so ohne, dieser Punkt. Denn: Algorithmen basieren nicht auf Regeln, sondern sie lernen aus Daten.

Von außen betrachtet erscheint die Entscheidungsfindung intransparent. Es ist für Dritte überhaupt nicht klar, woher die KI-Anwendung die Grundlage ihrer Entscheidungen nimmt. Was sind die echten Gründe, wenn nicht sogar die kausalen Gründe für die Vorhersage oder die Entscheidung?

Das ist aber nur ein wichtiges Kriterium. Es gibt auch andere, z. B. soziale oder auf die Umwelt bezogene.

Denken wir an Diversity, Gleichberechtigung und Fairness. Wir müssen dafür sorgen, dass KI nur mit solchen Daten gefüttert wird, welche keine Vorurteile (*bias*) aufweisen. Es gibt erschreckende Beispiele, etwa bei

der Gesichtserkennung (Differenzierung nach heller und dunkler Hautfarbe).

Datenschutz

Ganz besonders möchte ich noch auf den Datenschutz eingehen. KI, wenn sie denn auf Daten basiert, kann nur dann erfolgreich sein und auch wirklich funktionieren, wenn sie dem Datenschutz genügt.

Das Thema ist wichtig! Wer Daten besitzt, hat Macht. Seine eigenen persönlichen Daten zu schützen, bedeutet im Grunde nichts Anderes, als in unserer Gesellschaft frei zu sein. Selbstbestimmt kann nur der derjenige leben, der seine Daten unter Kontrolle hat – im Sinne von: Ich weiß, was mit meinen Daten geschieht.

Wollen wir höchstpersönliche Daten einfach blind in irgendwelche Hände geben, womöglich einer KI anvertrauen, die wir nicht verstehen? Wollen wir unsere wichtigsten persönlichen Informationen einfach so weitergeben?

Eines ist klar: Wir können heute die Frage nicht beantworten, ob und in welchen Bereichen KI irgendwann Menschen ablöst. Aber dass KI und Computer-Algorithmen irgendwann so gut sein werden, dass sie auch menschliche Emotionen, menschliche Gefühle, menschliche Bedürfnisse sehr, sehr präzi-



se und womöglich irgendwann präziser, als unsere Mitmenschen lesen können, steht für mich außer Zweifel.

Wenn wir damit umgehen wollen, dann geht es zuallererst darum, unsere höchstpersönlichen Daten zu kontrollieren.

Datenschutzlösung des AF Instituts

Das AF Institut hat hierfür eine Lösung, einen Rahmen gebaut, der diesen Umgang ganz spezifisch für die KI-Entwicklung mit Daten neu definiert. Wer einen Algorithmus für Maschinelles Lernen entwickeln möchte, der braucht dafür logischerweise Daten.

Die übliche Methode wäre, diese Daten zu nehmen, dem Analysten oder dem KI-Entwickler auf den Schreibtisch zu legen, dann werden die KI-Modelle entwickelt, getestet und irgendwann veröffentlicht. Dieser Prozess ist aus vielen Gründen gefährlich für den Datenschutz.

Beleuchten wir zunächst veröffentlichte KI-Modelle, also Modelle, die schon einen Trainingsprozess durchlaufen haben und jetzt im Einsatz sind. Sie haben aus den Daten gelernt! Die Information, die zunächst nur in diesen Daten steckte, ist jetzt im Modell.

Das Modell muss gar nicht mehr aktiv auf die Daten zugreifen, es ist bereits durch die Trainingsdaten geprägt. Es enthält diese Da-

ten. Wird es nicht vernünftig geschützt, dann werden sensible Informationen öffentlich zugänglich und können von Fachleuten abgefragt werden.

Daten, die in KI einfließen, müssen aus genau diesem Grund sicher sein. Wir müssen sicherstellen, dass diese Daten wirklich nichts verraten, was sie nicht verraten sollen.

Der herkömmliche Ansatz hierfür ist, Daten zu anonymisieren und personenbezogene Informationen zu streichen. Aber das reicht oft nicht. Es gibt viele und erschreckende Beispiele, die zeigen, dass auch aus scheinbar vollständig anonymisierten Daten durch Verknüpfung mit anderen Informationen tatsächlich Personen identifiziert werden konnten.

In Australien hat eine Forschergruppe bewiesen, dass aus anonymisierten Gesundheitsdaten der australischen Regierung Personen extrahiert werden konnten. D.h., die Forscher konnten sagen, welche Mitbürger tatsächlich hinter diesen vermeintlich anonymen öffentlichen Datensätzen stehen. Datenschutz ist somit weit komplexer, als viele glauben.

DQ0 – Datenschutz bei Open Access

Deswegen hat das AF Institute eine Methode bzw. einen Rahmen angewendet, in dem

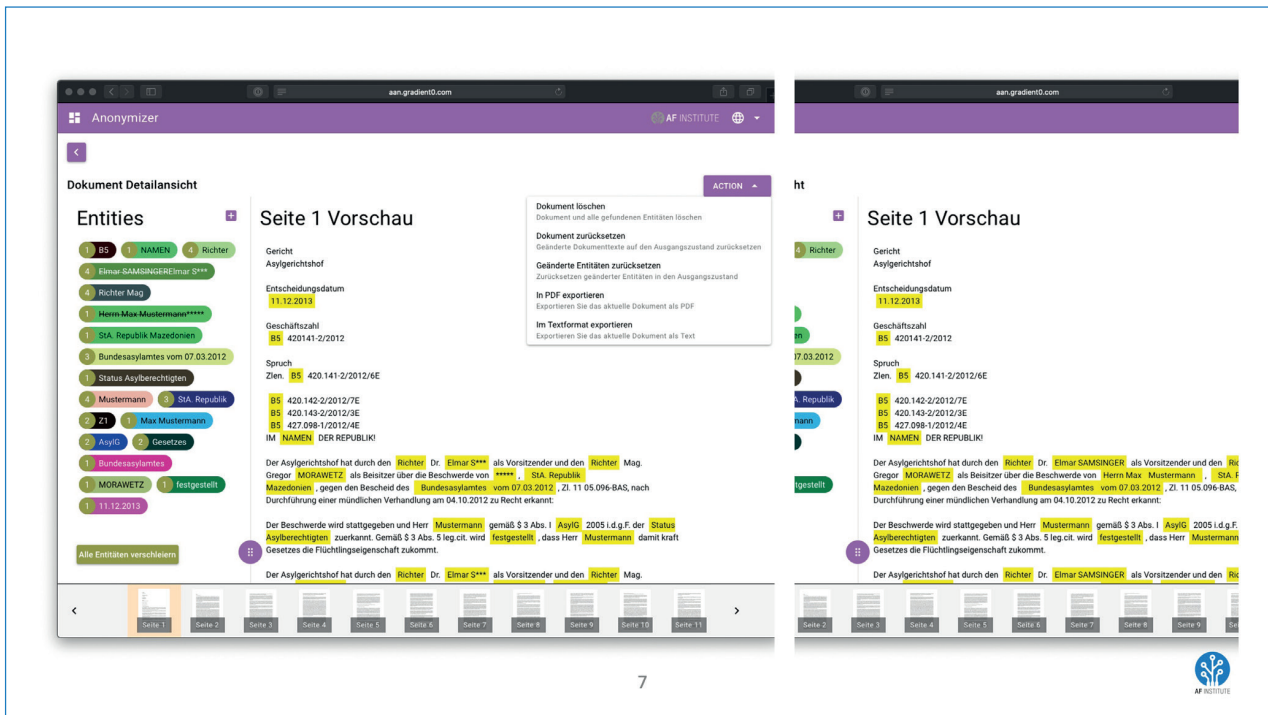


Abbildung 2

wir ein gängiges Prinzip quasi auf den Kopf stellen. Wir bringen nicht die Daten zur Analyse, sondern wir bringen die Analyse zu den Daten.

Die Daten bleiben, wo sie sind – sie wandern nicht an einen anderen Ort. Wir nennen diese Entwicklung *DQ0*, denn die sensiblen Informationen bleiben auf den Rechnern des Dateneigentümers, etwa einer Klinik, in Datenquarantäne.

Wenn jetzt eine Analyse, mathematische Methoden, Auswertungen oder beispielsweise eine KI-Algorithmik mit den Daten durchgeführt werden soll, dann „kommt die Analyse zu diesen Daten“. Die Operationen werden auf den Rechnern der Dateneigentümer durchgeführt. Die Maschine kann mit Daten höchster Güte operieren, da nichts pseudonymisiert oder anonymisiert wurde.

Aber: Wer immer die Operationen durchführt, sieht zwar das Ergebnis, kann aber die einzelnen, sensiblen Datensätze (z.B. den Namen und Wohnort eines Patienten) nicht abrufen. Auf diese Weise können z.B. Kliniken ihre Datenbestände bei Wahrung des Datenschutzes für Wissenschaftler öffnen (*open access*).

Die Datenoperationen werden sicher durchgeführt und mit von uns entwickelten

Zusatzmöglichkeiten kann dieser Sicherheitsstandard nochmals erhöht werden.

Differential privacy ist dann noch ein interessantes Thema. Diese Methode sorgt gemeinsam mit der Datenquarantäne dafür, dass Antworten, die hinausgehen, also die Ergebnisse der Analyse, so überprüft werden, dass sie datenschutzkonform erfolgen. Mit anderen Worten: Nichts wird verraten, was nicht verraten werden darf. Das passt gut zum Anspruch, den die allermeisten KI-Anwendungen ohnehin an sich selbst stellen.

Denn wenn wir statistische Analysen über Daten machen – und KI ist nichts Anderes als erweiterte Statistik –, dann möchten wir ja nicht wissen, welche Eigenschaften ein bestimmter Datenpunkt oder ein bestimmter Patient hat. Das interessiert eher den Therapeuten im konkreten Moment.

In der KI-Analyse wollen wir eher wissen, inwieweit Alter und Sterblichkeit bei Krebserkrankungen zusammenhängen, ob wir solche allgemeinen Informationen aus den Daten herausholen und der KI verfügbar machen können. Und dabei wollen wir die einzelnen Datenpunkte – Informationen über Menschen in diesem Fall – schützen. Wir müssen KI bei echtem Datenschutz einsetzen. Das schützt nicht nur sensible Personendaten, es

schafft auch Vertrauen für künftige Anwendungen der KI.

AFI Anonymizer – der totale Datenschutz

Für den Datenschutz hat das AF Institute noch eine weitere Option implementiert. Mit dem *AFI Anonymizer* (**Abb. 2**) können Datenbestände mit sensiblen Datensätzen veröffentlicht werden. Die Datenbestände ließen sich auf diese Weise sogar ins Netz stellen. Das ist also einer der wenigen wirklich vernünftigen Anwendungsfälle für Anonymisierung.

Wir haben beispielsweise eine Methodik entwickelt, die es erlaubt, sensible Dokumente so zu veröffentlichen, dass bestimmte persönliche Daten geschützt werden, andere hingegen nicht.

Innerhalb dieser – aus der Perspektive des Datenschutzes – sicheren Umgebung haben wir eine Lösung entwickelt, die Daten verarbeiten kann und dann über ML-Modelle und Sprachmodelle Entitäten (Namen, Orte, Begriffe usw.) erkennt und diese Entitäten im Dokument darstellt. Der Anwender kann diese Software benutzen, um diese Entitäten zu schwärzen.

Man sieht hier zwei wichtige Elemente: Zum einen erfolgt die Entwicklung in einem sicheren Rahmen. Diese *privacy by design* garantiert schon im Zuge der Entwicklung, dass die Datenschutzkriterien eingehalten werden. Gleichzeitig sorgen wir für *human oversight* –

der Mensch entscheidet am Ende, was veröffentlicht werden kann und was nicht. Der Mensch hat das letzte Wort. Nur er kann entscheiden kann, was tatsächlich richtig und schützenswert ist.

Wir haben den *AFI Anonymizer* ursprünglich für Gerichtsurteile entwickelt, um z. B. die Namen der Streitparteien zu löschen, aber die Namen der Richter beizubehalten. Damit können Richter sehen, wie Kollegen in ähnlich gelagerten Fällen geurteilt haben. Aber das Verfahren lässt sich genauso gut auf Bereiche im Gesundheitswesen anwenden, etwa Arztbriefe.

Conclusio

Zusammenfassend lässt sich sagen: KI bietet riesige Möglichkeiten. Denken wir nur an Textmodelle wie die maschinelle Auswertung von Arztbriefen. Genauso gut kann KI Datenbestände aus Krankenhaus-Informationssystemen auswerten – selbst bei unstrukturierten Daten lassen sich kontextspezifische Informationen ableiten.

Für die weitere Entwicklung von KI ist ein ethischer Ansatz von größter Bedeutung, sowohl hinsichtlich der KI-Freigabe und -Publikation also auch der Algorithmen.

Wenn wir diesen Grundsätzen der vertrauenswürdigen und vertrauensvollen ethischen KI folgen, dann wird KI einen entscheidenden Beitrag dabei leisten, das Leben vieler Menschen besser zu machen.

AF INSTITUTE
Working Paper