

## **WDR 5 Tiefenblick**

**WDR 5, Sonntag, 26.05.2019, 08:05 – 08:35 Uhr**

**Wiederholung: Sonntag, 26.05.2019, 22:30 – 23:00, WDR 5**

### **Mein Körper, meine Daten – Künstliche Intelligenz bedroht ärztliche Kompetenz (3/4)**

#### **Sprecher:**

Auf dem Campus der Hochschule Darmstadt. In einem großen Zelt, informieren Unternehmen über Berufschancen. Dichte Menschentrauben umlagern die Stände. Seit längerem bietet die Hochschule auch einen Studiengang für Datenwissenschaften an, Data Science.

„Praxisbezug“ heißt das Motto. Wir geben Unternehmen immer wieder mal Gelegenheit, sich bei uns zu präsentieren, sagt mir der Mathematikprofessor Sebastian Döhler. Ich besuche ihn, weil ich wissen will, wo genau die Grenzen der Big-Data-Methoden im Gesundheitswesen liegen.

#### **O-Ton Döhler:**

Unsere guten Absolventen gingen schon immer weg wie warme Semmeln, muss man sagen. Weil Statistik schon immer gebraucht wurde. Was glaube ich, schon neu ist, und das ist sicherlich eine Folge der Digitalisierung, dass diese Daten jetzt gesammelt werden und dass sie nicht mehr als nutzloser Ballast empfunden werden, sondern dass man den Eindruck hat, wir können etwas damit machen. Oder wir sollten etwas damit machen.

#### **Sprecher:**

Was machen gut ausgebildete Data-Wissenschaftler im Gesundheitswesen? Werden sie tatsächlich – wie manche behaupten – Medizin und Psychiatrie revolutionieren? Wie könnte ein sinnvoller Umgang mit Patientendaten aussehen?

**Titelansage:**

Das ist Tiefenblick mit der Reihe: Mein Körper, meine Daten, Teil 3: Künstliche Intelligenz bedroht ärztliche Kompetenz. Ich bin Martin Hubert.

**Sprecher:**

Bei meiner Recherche stoße ich auf eine Unmenge von Versprechungen aus dem Big-Data-Bereich:

Das Houston Methodist Research Institute vermeldet, dass ein selbstlernendes maschinelles System Mammografieberichte von Brustkebspatientinnen 500 mal schneller auswertet als Ärzte. Dabei werde Krebs zu 99 Prozent richtig erkannt, von Ärzten nur zu 78 Prozent.

Das am New Yorker Mount Sinai Klinikum entstandene System „Deep Patient“ soll eigenständig Muster in Patientendaten finden und so Prognosen für eine Vielzahl an Krankheiten liefern.

Das Unternehmen AliveCor verkündet, dass es die Apple Watch mit einem neuronalen Netz ausgestattet hat, welches die Herzfrequenz der Benutzer nicht nur überwacht, sondern sogar ein EKG ableiten kann.

Eine Software der Firma Heart Flow verspricht den Grad der Verkalkung von Herzkranzgefäßen aufgrund von CT- Scans automatisch analysieren zu können.

Es herrscht Goldgräberstimmung und so habe ich an der Hochschule Darmstadt, wo Studierende die neuesten Methoden der Data Science lernen, eine euphorische Stimmung erwartet. Aber der Grundtenor ist ein anderer.

**O-Ton Döhring:**

Hier vermitteln wir den Studierenden auch den kritischen Umgang mit Ergebnissen. Also nur weil man eine bestimmte Methode verwendet hat, die momentan vielleicht sehr stark präsent ist, heißt das noch lange nicht, dass, dass man damit gute Ergebnisse erzielen kann. Es kommt immer auf den Einzelfall hat.

**Sprecher:**

Professor Markus Döhring koordiniert den Studiengang gemeinsam mit Sebastian Döhler.

**O-Ton Döhring:**

Vielleicht kennen sie früher noch diese verschwommenen Bildchen, wo man lange genug draufgestarrt hat und irgendwann sprang dann so ein 3-D Effekt entgegen. Und genauso ist es mit manchen Datensätzen auch, wenn man nur lange genug draufstarrt, dann springen einem irgendwann Dinge entgegen, die vielleicht auffällig sind, aber die gar keine validen Schlüsse sind.

**O-Ton Malcherek:**

Das ist immer der Wunschtraum: ich nehme bestehende Tools, stecke die zusammen wie so einen Legobaukasten und kriege meine Antworten.

**Sprecher:**

Professor Arnim Malcherek, der den Darmstädter Studiengang mit aufgebaut hat, stößt ins gleiche Horn.

**O-Ton Malcherek:**

In der Realität liegen die Daten leider nie in der Form vor, wie man sie braucht, d.h. man muss sie wandeln, säubern, irgendwo reinfüttern, aufbereiten noch mal, indem man die Tools kombiniert, die es schon gibt. Und die eine Blackbox, wo ich meine Daten rein stecke und dann kriege ich eine Antwort raus – funktioniert nach meiner Erfahrung auf einfachen Fragestellungen überraschend häufig gut, auf komplexeren Fragestellungen erwartet schlecht.

**Sprecher:**

Die selbstkritischen Antworten in Darmstadt haben mich positiv überrascht. Sollten diejenigen, die sich genauer mit den Methoden des Big-Data-Booms auseinandersetzen, gar nicht so sehr an die großartigen Versprechungen glauben, die mit ihnen verbunden sind? In Freiburg treffe ich mich mit einem der schärfsten Kritiker von Big Data im Gesundheitswesen.

3/17

**O-Ton Antes:**

Das Ganze ist eine Ideologie, ohne jeden Zweifel.

**Sprecher:**

Professor Gerd Antes ist Mathematiker und Biometriker und war bis Oktober 2018 Direktor des deutschen Cochrane Zentrums in Freiburg. Die internationale Cochrane Organisation kämpft für eine so genannte evidenzbasierte Medizin. Eine Medizin, die auf nachvollziehbaren Methoden und exakt belegten Ergebnissen beruht. Auch nach seiner Tätigkeit bei Cochrane setzt Antes diesen Kampf fort und hat besonders Big Data im Visier.

Schon die so genannte Präzisionsmedizin, die individuelle Varianten von Tumoren bekämpfen will, sagt Antes, sei mit Vorsicht zu genießen. Denn normalerweise gelten Behandlungsmethoden oder Medikamente nur dann als wirksam, wenn sie bei vielen Menschen gleichermaßen wirken. Deshalb werden Methoden und Medikamente in Studien auch an vielen Menschen geprüft, damit sie auch bei vielen Menschen erfolgreich eingesetzt werden können. In der Präzisionsmedizin dagegen werden viele Daten nur analysiert, um allein die individuelle Ausprägung einer Krankheit eines einzelnen oder weniger Patienten zu behandeln.

**O-Ton Antes:**

Da gibt es direkt den Begriff des Paradoxons der Präzisionsmedizin, weil sie je kleiner sie die Einheit wählen und dann wirklich nur noch ein Individuum haben, keine Chance mehr haben zu prüfen, ob das, was sie tun, richtig ist. Also ob das tatsächlich wiederholbar ist bei klinischen Studien. Bei Wissen ist ja immer die Frage: ist es übertragbar? Und je feiner die Einheit ist, desto weniger ist klar, ob es übertragbar ist.

**Sprecher:**

Kann das dem einzelnen Patienten nicht egal sein, frage ich? Wenn die auf ihn zugeschnittene Behandlung hilft, wird er doch auf die ehernen Regeln evidenzbasierter Wissenschaft pfeifen. Nicht jede individuelle Behandlung funktioniert, sagt Antes, ihr Erfolg sei schwer voraussagbar. Und ein Tumor, für den man eine passende Wirkstoffkombination gefunden hat, könne sich verändern – dann stehe man wieder am Anfang.

Kann es Fortschritt in der Medizin überhaupt geben, ohne dass Wissenschaftler etwas wagen und dabei Risiken eingehen? Manche Ärzte glauben, dass neue Studienverfahren entwickelbar sind, mit denen sich die Ergebnisse der Präzisionsmedizin überprüfen lassen. Wenn solche gefunden werden und sie gut sind, sagt Antes, werde er keine Einwände erheben. Aber es sei ein Irrglaube zu meinen, dass es bei den Versprechungen des medizinischen Fortschritts immer nur um Gesundheitsinteressen ginge.

**O-Ton Antes:**

Das ökonomische Interesse ist an vielen Stellen größer als das medizinische Interesse.

**Sprecher:**

Nicht nur Gesundheitsunternehmen, auch Wissenschaftler stehen unter Konkurrenzdruck. Sie müssen hart arbeiten, um Forschungsgelder einzutreiben und viele Studien veröffentlichen, um Karriere machen zu können. Vieles sei daher PR oder wissenschaftliche Übertreibung, sagt Antes.

**O-Ton Antes:**

Wir haben das ja auch vorher schon gehabt! Wir haben ja immer gehört, wenn das menschliche Genom entschlüsselt ist, dass wir dann wissen: Da ist das Symptom, dann ist da die Stelle im Genom, an der ich etwas ändern muss. Und dann hat man irgendwann gemerkt, dass es nicht eine Stelle ist, sondern vielleicht tausende oder zwanzigtausend und wenn dann die auch noch alle in Wechselwirkung zueinander stehen, dann bin ich natürlich in einer extrem schwierigen Situation und dann ist genau das, was versprochen wurde, nicht erreichbar.

**Sprecher:**

Und wie sieht es aus mit selbst lernenden Systemen? Sie werden mit einer riesigen Datenmenge trainiert, zum Beispiel mit Scans diverser Tumore, EKGs von Herzrhythmusstörungen oder Röntgenaufnahmen von Lungendefekten. Eine Künstliche Intelligenz entwickelt dann ein Muster, mit dem sie auch bei neuen Patienten krankhafte Veränderungen entdecken könnte. Ist das kein Segen?

Man müsse die Qualität der einzelnen Verfahren testen und dann werde es auch sinnvolle Anwendungen geben, gesteht Antes zu. Aber Wunderapparate seien das nicht.

**O-Ton Antes:**

Das ist glaube ich die größte Schwäche von diesen Maschinen, dass sie alle nicht richtig intelligent sind, sondern tatsächlich das machen, was sie machen sollen, vielleicht gut. Aber das, was der Arzt zum Beispiel macht, zum Beispiel der Hautarzt, vor dem man sich ganz auszieht, dann nach irgendetwas schaut und sagt, „Hei, was ist denn da unten los“, und dann Dinge sieht – schlimmsten- oder bestenfalls eine Krebserkrankung – auf die er eigentlich sonst nicht gekommen wäre und weswegen der Patient auch nicht dort war. Wenn jetzt die Maschine genau fokussiert etwas macht und nicht links und rechts guckt, weil sie es gar nicht kann, dann werden die Dinge übersehen und diese Gefahr, die wird man datenmäßig kaum in den Griff kriegen, weil man ja nicht sieht, was die Maschine hätte sehen können.

**Sprecher:**

Trotz dieser Schwäche, sagen die Befürworter, könnten KI-Systeme immer besser werden, wenn man sie mit immer *mehr* Daten füttert. Ein System, das viele Bilder von etwas gesehen hat, besitze ein schärferes und differenzierteres Muster davon, könne es also besser erkennen. Auch so ein Irrtum, sagt Antes.

**O-Ton Antes:**

Google Flu ist eines der Beispiele, die auf teilweise sehr hämische Weise in einer der großen Zeitschriften besprochen wurden.

**Sprecher:**

Bereits 2009 wollten Wissenschaftler von Alphabet, damals noch Google, ein mustergültiges Big-Data-Produkt erzeugen: „Google Flu“. Fünf Jahre lang folgten sie amerikanischen Staatsbürgern und Millionen von Suchbegriffen, die diese bei ihrer Internetsuche eingaben: Schnupfen, Fieber, Kopfschmerz und so weiter. Sie fanden 45 davon und brachten sie mit Statistiken realer Grippeausbrüche in bestimmten Regionen zusammen. Um zukünftige Grippeepidemien vorherzusagen, damit das Gesundheitssystem sich darauf einstellen kann, erzeugten sie einen Algorithmus, der die Google-Suchbegriffe nach dem Motto überwachte: Nimmt das Begriffscluster stark zu, steuern wir auf eine Grippewelle hin.

**O-Ton Antes:**

Dass das nicht unbedingt klappen muss, das war eigentlich von vornherein klar, weil natürlich auch durch Gerüchte, durch Medienberichte die Klickanzahl erhöht werden kann, und nicht nur weil jemand krank ist, sondern auch nur vielleicht, weil jemand aufgeschreckt wurde. Deswegen fand ich es überhaupt nicht überraschend, dass das nicht gut geklappt hat. Aber ist dann wirklich schief gegangen nach einiger Zeit. Und dann kam von denjenigen, die das Prinzip retten wollten, sofort die Aussage, ja es war auch Google schuld, weil die sofort aufgehört haben, das zu verfolgen, die hätten immer weiter machen sollen. Weil „mehr Daten“ ja alles löst. Aber dumm ist Google bestimmt nicht. Und die haben einfach gesehen dass es ein Irrweg war, und haben es eingestellt.

**Sprecher:**

Beispiel zwei: Watson, die viel gerühmte künstliche Intelligenz von IBM. Sie wurde mit Daten von Krebspatienten trainiert und sollte Krebs besser diagnostizieren können als Tumorspezialisten.

**O-Ton Antes:**

Da gibt es das berühmte Beispiel von einer der größten Krebskliniken auf der Welt, in Houston. Und die haben Doktor Watson entlassen, nachdem sie 50 Millionen Dollar investiert hatten, weil sie festgestellt haben, dass er nicht besser ist als die Ärzte.

**Sprecher:**

IBM hatte Watson mit Daten von sehr schweren Krebserkrankungen trainiert. Die ließen sich aber nur schwer auf andere Patientenkollektive übertragen.

Solche Beispiele widerlegen die Behauptung, dass es vor allem auf „Big Data“, also auf möglichst viele Daten ankommt. Kritiker sprechen von einer „blinden Datenanalyse“, wenn man sich hauptsächlich auf die Anzahl der Daten verlässt und KI-Programme quasi blind nach Mustern suchen lässt.

Allerdings sind Gegenbeispiele auch nur Beispiele und es ist nicht zu leugnen, dass auch erfolgreiche Projekte von Datenverarbeitung Künstlicher Intelligenzen existieren.

Was zeichnet die aus? Und wie beurteilt jemand die Einwände von Gerd Antes, dem ein solches Projekt gelungen ist?

**O-Ton Lengauer:**

Erstens hat Herr Antes natürlich Recht: Blinde Datenanalyse ist in vielen Fällen nicht zielführend und kann auch sehr irreführend sein. Aber, das ist auch unsere Erfahrung, man kann, wenn man ein gutes Datenmodell hat, dass genau Vorhersagen trifft, kann man wirklich sehr viel bewirken, ohne genau zu wissen, was die unterliegenden Mechanismen sind.

**Sprecher:**

Professor Thomas Lengauer war bis 2018 Direktor am Max-Planck-Institut für Informatik in Saarbrücken und ist Präsident der Internationalen Gesellschaft für Bioinformatik. Seine Arbeitsgruppe forscht seit vielen Jahren daran, die medikamentöse Behandlung von Patienten mit HIV zu verbessern. Keine einfache Sache, erklärt er mir in meinem Büro, denn HI-Viren werden ständig gegen Wirkstoffe resistent, sodass man die Medikamente intelligent kombinieren muss. Um das Kombinieren zu erleichtern, hat Lengauers Team Daten über Wirkstoffe und ihre Effekte auf das Virus in ein KI-System eingegeben.



**O-Ton Lengauer:**

Wir haben also Vorhersagen für jeden einzelnen dieser Wirkstoffe und der Server gibt eine Liste zurück, die sagt, was das geschätzte Resistenzniveau gegen jeden dieser Wirkstoffe ist. Und diese Liste nimmt sich der Arzt und überlegt sich jetzt, wie er Wirkstoffe kombinieren will. Dieser zweite Schritt ist immer noch in der Version manuell vom Arzt zu tun.

**Sprecher:**

Das KI-System sei erfolgreich und großflächig im Einsatz, versichert Lengauer. Es spuckt sogar Informationen darüber aus, aus welchem Grund ein Virus resistent gegenüber einem Wirkstoff ist.

Warum – frage ich ihn – kann das System nicht automatisch auch die Entscheidung über die Medikamentenkombination treffen? Warum macht das noch der Arzt?

Sie werden überrascht sein, sagt Lengauer, aber seit vielen Jahren habe sein Team eine zweite Form dieses Systems entwickelt, das genau das könne. Aber es wurde aus gutem Grund von den Ärzten nicht angenommen.

**O-Ton Lengauer:**

Der Grund ist, dass die erste Form bereits umfangreicheren Studien unterworfen wurde, wo andere Leute getestet haben, ob das wirklich funktioniert. Solche Studien sind für die zweite Form bisher nicht gemacht worden und aus dem Grunde ist die zweite Form in dem Sinne nicht so gründlich validiert und Mediziner sind sehr konservativ, wenn es nicht validiert ist, dann wollen die das nicht machen. Und das finde ich auch richtig so.

**Sprecher:**

Klingt für mich vorbildlich: Entwickele intelligente datenverarbeitende Systeme, die dem medizinischen Fortschritt dienen können. Setze sie aber nur dann ein, wenn ihre Wirkung in guten Studien validiert, also bestätigt wurde. Zum Beispiel in so genannten prospektiven Studien. Die dienen etwa der Überprüfung einer Hypothese *zur Wirksamkeit* einer Behandlungsmethode an gezielt dafür zusammengestellten Daten von Versuchspersonen.

Hat Thomas Lengauer die erste Variante seines Systems mit einer prospektiven Studie getestet?

9/17

**O-Ton Lengauer:**

Das ist bei uns nicht gemacht worden. Erstens sind prospektive Studien extrem teuer und zweitens ist es tatsächlich auch so, dass zumindest bei den prospektiven Studien klassischer Prägung, die meistens benutzt werden, wir sehr niedrigdimensionale Daten haben.

**Sprecher:**

Das heißt: Die Datenmenge ist im Vergleich zu Big-Data-Studien gering.

**O-Ton Lengauer:**

Also diese prospektiven Studien, die können mit hochdimensionalen Daten aus meiner Sicht überhaupt noch nicht umgehen. Mit anderen Worten: Wir haben jetzt Methoden entwickelt – aber die Testmethodik dafür, die gibt es noch gar nicht, ja, (*lacht*), so richtig.

Das wird auch von den Regulierungsbehörden erkannt, die normalerweise prospektive Studien verlangen. Bevor überhaupt ein Wirkstoff auf den Markt kommt, müssen prospektive Studien gemacht werden, diese teuren Dinger, darum dauert es auch so lange, bis ein Wirkstoff auf dem Markt ist. Aber in unserem Bereich hat die Regulierungsbehörde gesagt „ Uuh, ich weiß auch nicht wie das geht und das ist zu kompliziert, macht ihr mal eure retrospektiven Studien, wir halten da still“. Das ist für den ganzen HIV-Bereich, den gesamten HIV-Bereich der Fall.

**Sprecher:**

Retrospektive Studien können die Wirksamkeit eines Verfahrens nicht so gezielt überprüfen wie aufwändige prospektive Studien. Sie arbeiten mit alten, bereits vorhandenen Daten, die nicht gezielt für dieses spezielle Verfahren ausgewählt wurden.

Der medizinische Fortschritt ist also eine heikle Angelegenheit, gerade auch, wenn es um selbstlernende Systeme und große Datenmengen geht.

Befürchtet Thomas Lengauer nicht, dass er die Grenzen seriöser Wissenschaft überschritten hat, wenn auch mit Duldung der Regulierungsbehörden?

Nicht bei unserem System, sagt er, denn es sei wirklich mehrere Jahre lang an mehr als einem Dutzend retrospektiver Studien getestet worden. Und es werde verantwortungsbewusst eingesetzt.

10/17

**O-Ton Lengauer:**

Natürlich muss man bei der Software sagen: Leute, wir haben das im Durchschnitt getestet und es geht ganz gut, aber im Einzelfall kann es immer noch total falsch sein. Das muss man oben drüber schreiben, das schreiben wir auch bei unserem Server oben drüber. Und dann liegt die Verantwortung beim Arzt. Und so muss KI meiner Ansicht nach auch geschneidert sein, dass letztlich die Technik die Verantwortung nicht abnimmt. Das ist eben die Frage: Wollen wir KI als Assistenten haben, als kognitive Lupe oder wollen wir sie als Master haben, die autonom etwas viel besser macht als wir und da bin ich eindeutig bei Variante Nummer eins mit sehr vielen Leuten auf der Welt. Es gibt auch genug Leute, die Variante Nummer zwei verfolgen, auf die müssen wir richtig aufpassen.

**Sprecher:**

Zwei Fraktionen von Forschern konkurrieren um die zukünftige Ausrichtung der Big-Data-Medizin. Die Vorsichtigen, die maschinelle Datenanalyse als Instrument sehen und als Assistenz des Arztes im Dienste des Patienten. Und – ich nenne sie mal – die Radikalen, die der Technik Autonomie wünschen und ihr Wunderdinge zutrauen.

Was ist zu tun, damit der Mensch Herr der Entwicklung bleibt? Diese Frage stellt sich zurzeit vor allem auch auf dem Gebiet der Psychiatrie. Denn hier beginnen Wissenschaftler gerade damit, die Big-Data-Forschung zu intensivieren. Und hier geht es um das Innerste des Menschen, seine Psyche und seinen Geist.

**O-Ton Walter:**

So einfach ist die Sache nicht.

**Sprecher:**

Professor Henrik Walter treffe ich in seiner Wohnung in Berlin-Mitte. Er ist Psychiater, Hirnforscher und Philosoph und forscht mit vielen Mitarbeitern an der Charité zu unterschiedlichen Themen. Warum interessiert er sich für den Einsatz intelligenter datenverarbeitender Systeme, wenn die Sache nicht einfach ist?

Der Grund dafür sind Studien, die vorgeben, Gesunde und psychisch Erkrankte viel genauer als bisher unterscheiden zu können, erklärt er mir. Für die Diagnose wäre das ein Meilenstein, aber die Studienergebnisse müssen erst noch genau überprüft werden.

Im einfachsten Fall untersucht man dafür Unterschiede in den Hirnscans von Hirnarealen depressiver und gesunder Versuchspersonen.

**O-Ton Walter:**

Wenn man dort eine Genauigkeit von 50 Prozent hat, dann ist es ja wie Zufall – in der Literatur werden Werte von 80, 90 Prozent berichtet. Und jetzt gibt es einen geschätzten Kollegen, den Herrn Tim Hahn, der hat eine sogenannte „peck“ gemacht, das ist so eine Art competition, wo es einen Datensatz gibt, und die gesagt haben, schaut mal, wir geben euch allen die gleichen Daten und mal gucken wer das am besten unterscheiden kann.

**Sprecher:**

Das Besondere an diesem Wettbewerb: Anders als bei bisherigen Studien über Merkmale der Hirnscans psychiatrisch erkrankter Menschen wurde nicht an kleinen Stichproben getestet, sondern an einem der größten Datensätze von Studien in diesem Bereich.

**O-Ton Walter:**

Wir haben auch gemeinsam mit einer jungen Juniorprofessorin für Psychiatrie Kerstin Ritter das gemacht und die neusten Methoden eingesetzt und kamen einfach nicht auf eine bessere Genauigkeit als ca. 65 Prozent, 67 Prozent. Na gut, vielleicht ist es halt so, haben wir gedacht, schicken wir das mal ein. Wir haben dann den dritten Platz gemacht und als das Ergebnis jetzt veröffentlicht wurde, haben wir gesehen, dass auch die anderen nur ein ganz bisschen besser sind, keine ist über 70 Prozent gekommen. Warum? Denn 70 Prozent ist nicht toll, das ist noch immer eine Fehlerwahrscheinlichkeit von 30 Prozent.

**Sprecher:**

Walter meint, dass die eingesetzten Algorithmen möglicherweise nicht gut genug gewesen sind. Aber zeigt dieser Wettbewerb zum Maschinenlernen in der Psychiatrie nicht zuallererst, dass bisher viel zu viel versprochen wurde?

**O-Ton Walter:**

Sie wissen ja, wie das ist, das ist wie in der Wirtschaft ein Hype, da muss jetzt jeder auf den Zug draufspringen und man selber muss extrem aufpassen, dass man selber keinen Mist produziert. Denn es gibt nur wenige Leute, die von allen diesen Sachen etwas verstehen.

**O-Ton Hahn:**

Dieser Hype muss irgendwann wieder mal runterkommen und aktuell ist es relativ leicht mit bestimmten Schlüsselworten eine Menge Geld einzuwerben.

**Sprecher:**

Tim Hahn ist der Initiator des Wettbewerbs, an dem auch der Psychiater Henrik Walter teilnahm, um Gesunde und psychisch Kranke mit Hilfe einer Künstlichen Intelligenz zu unterscheiden. Ich treffe Hahn in seinem Büro am Fachbereich Medizin der Universität Münster. Hahn leitet hier die Arbeitsgruppe „Künstliche Intelligenz in der Psychiatrie“. Er ist ein junger Wissenschaftler, der das Gebiet maschineller Datenverarbeitung in Deutschland voranbringen will. In ersten Studien konnte er zum Beispiel vorhersagen, welche Patienten mit Panikattacken von einer kognitiven Verhaltenstherapie profitieren und welche nicht.

**O-Ton Hahn:**

Wir haben eine ganz ähnliche Studie gemacht, da geht es um EKT, also um die Elektrokrampftherapie, die eine der am besten nachgewiesenen Therapien bei chronischen Depressionen ist, die aber natürlich andererseits auch ein Verfahren darstellt, das für den Patienten eine sehr belastende Intervention sein kann.

**Sprecher:**

Hahn arbeitet daran, aus Daten von Smartphones und anderen tragbaren Geräten vorherzusagen, ob jemand psychisch krank wird. Trägt er mit solchen Studien nicht mit zu dem Hype bei, den er eigentlich kritisch sieht?

**O-Ton Hahn:**

13/17

Also das muss man einschränkend feststellen, es handelt sich um Proof of Concept Studien.

**Sprecher:**

Damit meint er den allerersten Schritt zu einer klinischen Studie, der nur testet, ob ein erdachtes Konzept überhaupt funktionieren kann.

**O-Ton Hahn:**

In diesen Studien sind in der Regel zwischen 60 und einige 100 Leute. Ob die in der klinischen Anwendung in den nächsten Jahren dann tatsächlich Fuß fassen können, das wird davon abhängen, wie weit das in der Fläche bei einer großen Anzahl von Patienten funktioniert.

**Sprecher:**

Dazu sind dann weitere Studien mit viel mehr Versuchspersonen nötig. Wie aber will Hahn den Spagat schaffen, einerseits vor dem Hype und vorschnellen Versprechungen zu warnen und andererseits Pionier zu sein?

**O-Ton Hahn:**

Ich glaube, dass wir uns auf die Füße fallen, wenn wir nicht rigorose Qualitätskriterien einziehen. Und da wünsche ich mir, dass für machine learning-Modelle, die wir tatsächlich in der klinischen Praxis haben wollen, solche Kriterien etabliert werden.

**Sprecher:**

Er selbst habe drei solcher Kriterien oder Regeln ausgearbeitet.

**O-Ton Hahn:**

Der erste Schritt: Wir gucken uns den zusätzlichen Nutzen des Modelleinsatzes an.

**Sprecher:**

14/17

Radikale Big-Data-Befürworter behaupten: Ein selbst lernendes maschinelles System, das gut funktioniert, wird besser sein als der Mensch.

Tim Hahns erste Regel besagt: Das ist keineswegs ausgemacht, denn auch gut funktionierende Big-Data-Systeme machen Fehler. Diese Fehlermuster müsse man dann vergleichen mit solchen Fehlern, die Ärzte oder Psychiater machen. Dann erst könne man entscheiden, ob das KI-System wirklich mehr Nutzen bringt.

Regel Nummer zwei: Prüfe genau, ob die Ergebnisse eines Systems verallgemeinerbar sind! Radikale Big-Data-Befürworter behaupten: Je mehr Daten wir in ein System eingeben, desto besser funktioniert es. Tim Hahn widerspricht.

**O-Ton Hahn:**

Da spielt es erst einmal keine Rolle, an wie vielen Leuten ich die trainiert habe, ob ich da jetzt 50 Leute drin habe oder 500 oder 50.000 oder 5 Millionen. Der wirklich springende Punkt ist. Wenn ich die an einem großen Datensatz teste, funktioniert er dann immer noch?

**Sprecher:**

Entscheidend ist also nicht die Eingabe von möglichst vielen Daten, damit das System lernt, sondern ob das System an vielen neuen Daten nachweisbar funktioniert. Um solche Tests zu unterstützen, hat Tim Hahn mit Mitstreitern ein Online-Portal gegründet, über das Forschergruppen ihre Verfahren gegenseitig überprüfen können – freiwillig.

Dritte Regel: Setzt Big-Data-Systeme differenziert ein!

Radikale Big-Data-Befürworter träumen von möglichst universellen Systemen. Tim Hahn pocht darauf, dass transparent sein müsse, mit welchen Daten ein System trainiert wurde, damit man es nur bei Patientengruppen mit denselben Charakteristika einsetzt. Andernfalls besteht die Gefahr, dass das System versage, wie das IBM-System „Watson“, das an Daten schwerer Tumorerkrankungen trainiert wurde und bei leichteren Formen scheiterte.

**O-Ton Hahn:**

15/17

Wir möchten ganz viele Modelle haben, aus denen ein System, das den Kliniker unterstützt, letztendlich dasjenige auswählt, was zu den Charakteristika dieser Person passt. Und dafür das Modell auswählt, das hier die geringste Fehlerquote hat. Und das hier für den Zweck, für den es trainiert wurde, optimal angewendet werden kann.

**Sprecher:**

Drei Regeln, die nicht nur für die Psychiatrie relevant sind, sondern für die gesamte Medizin. Hahns Bemühungen demonstrieren, dass an einer überprüfbaren und realistischen Variante der Big-Data-Medizin gearbeitet wird.

Ganz gleich wie der Kampf zwischen radikaler und realistischer Big-Data-Medizin ausgehen wird – Daten von Patienten werden in Zukunft stärker fließen: für die Forschung und die maschinellen Systeme.

Welche Folgen hat das für den Datenschutz? Welche ethischen Fragen wirft das auf?

Halten Informatiker wie Markus Döhring, der an der Hochschule Darmstadt den Data-Science Studiengang mitkoordiniert, solche Fragen überhaupt für relevant?

**O-Ton Döhring:**

Halte ich für sehr relevant, weil viele – und da nehme ich mich gar nicht aus – viele technikverliebte Personen natürlich erst mal zufrieden sind, wenn die Lösungen entwickelt sind, aber manchmal nicht ausreichend über die weiteren Folgen nachdenken. Und dafür zu sensibilisieren, halte ich für eine Verantwortung von uns der Gesellschaft gegenüber, die Studenten dafür fit zu machen für solche Fragestellungen, die aus meiner Sicht in der Zukunft sehr stark zunehmen werden, weil sie momentan noch verdeckt oder für die allgemeine Bevölkerung gar nicht sichtbar genug sind.

**Titelabsage:**

Das war Tiefenblick mit der Reihe: Mein Körper, meine Daten. Teil 3: Künstliche Intelligenz bedroht ärztliche Kompetenz, von Martin Hubert.



**Sprecher:**

Im vierten und letzten Teil geht es um die Frage: Kann es einen ethischen und verantwortungsbewussten Einsatz von Big-Data-Technologie überhaupt geben? Im medizinischen Alltag, in der Forschung oder im Versicherungswesen?

**O-Ton Woopen:**

Man kann nicht so tun, als sei ein Algorithmus von vornherein ethisch neutral. Ganz im Gegenteil! Wir müssen sehr gut darauf achten, was mit diesen Daten, die gesammelt werden, getan wird, wo es getan wird und von wem es getan wird.

**Sprecher:**

Das alles nächste Woche in Tiefenblick auf WDR 5 – oder alle Folgen im Podcast , im WDR Featuredepot oder der ARD Audiothek.

Eine Produktion des Westdeutschen Rundfunks 2019.