

Transkript*: Metaverse meets Health 2026: Keynote von Prof. Dr. Ekaterina Jussupow

*Bitte beachten: Das Transkript wurde automatisiert erzeugt und wurde nicht nachträglich gegengelesen oder korrigiert. Abweichungen vom Wortlaut können daher nicht ausgeschlossen werden. Bei Rückfragen wenden Sie sich bitte an: inklusive-videos@fernuni-hagen.de

(*bei Rückfragen: inklusive-videos@fernuni-hagen.de*) Wir kommen zur zweiten Keynote jetzt von der Kollegin Jusupow. Mein Namensschild fällt schon von Anfang an ab. Ich glaube, die beiden Keynotes bauen recht gut aufeinander auf. Wir hatten eben gehört, wie es ist zu fliegen oder operiert zu werden von Menschen, die noch nicht trainiert sind. Jetzt stellt sich die Frage, lassen wir uns fliegen oder operieren von einer künstlichen Intelligenz? Ich glaube, vielleicht der eine oder andere investiert mit KI-Advisern. Möglicherweise ist vielleicht die Hemmschwelle nicht ganz so hoch, aber wenn jemand dann im Flugzeug KI-gesteuert fliegt oder in der Medizin KI-gesteuert mich operiert, da wäre ich zumindest noch ein bisschen vorsichtiger. Ich bin eigentlich auf der nerdigen Seite des Lebens als Wirtschaftsinformatiker. Wir sind froh, dass wir nach dem medizinischen Vortrag jetzt eine Referentin aus der Wirtschaftsinformatik haben. Frau Professorin Jusupow ist seit März 2023 Professorin an der TU Darmstadt in Wirtschaftsinformatik, also eine Fachkollegin, deswegen konnten wir sie auch gewinnen. Deine Forschungsschwerpunkte sind Mensch-KI-Interaktion. Ich glaube, dazu hören wir gleich auch ein bisschen was jetzt hier in unserem medizinischen Entscheidungskontext. Und du bist auch Psychologin, von daher passt das, glaube ich, von deinem Profil her sehr gut in die heutige Veranstaltung. Dann möchte ich an dich übergeben. Viel Spaß bei deiner Keynote. Ich darf dann im Anschluss auch die Fragerunde moderieren und danach geht es dann weiter in den noch interaktiveren Teil, im hinteren Teil des Raums. Also gutes Gelingen. Dankeschön. (...) Ich habe jetzt die Freude, nach der sehr praktischen Anwendung, wo wir wirklich die Realität gesehen haben, mal ein bisschen mehr in die Theorie zu gehen. Tatsächlich werde ich Ihnen ganz viele Sachen mitbringen, die sehr tief in menschliche Kognitionsprozesse reingehen. Also Sie werden ganz viele

verschiedene Sachen zum Thema menschliches Denken kennenlernen. Wie Sie das dann in die Praxis übersetzen können, das würde ich es Ihnen überlassen. Wir können sehr, sehr viele Sachen auch diskutieren, wie KI-Systeme vor allem in der medizinischen Praxis auch angesetzt werden. Aber ich habe bewusst meinen Vortrag jetzt viel mehr auf diese Forschungsergebnisse gelegt. Nicht, weil ich davon nicht so viel weiß, sondern eher einfach nur, um ein bisschen einen Akzent zu setzen. Was meine ich denn, wenn ich von KI-Systemen spreche? Wir unterscheiden verschiedene Arten von KI-Systemen. Ein traditioneller würde ich jetzt nicht sagen, aber so ein System, was sehr, sehr viel Anwendung vor allem in der medizinischen Diagnostik, vor allem in der Radiologie findet, sind sogenannte prädiktive Algorithmen, die in der Lage sind, zum Beispiel verschiedene Bilder zu klassifizieren. Also diese Modelle funktionieren so, dass da mehr oder minder verschiedene neuronale Netze in verschiedenen Layers zusammengelegt werden, um dann praktisch pixelbasiert eine Klassifikationsentscheidung zu treffen. Dann steht am Ende zum Beispiel hier so ein Bild, das geht hier darum, zu zeigen, wo ist eigentlich jetzt ein Schlaganfall und in welchen Gehirnregionen ist er dann eigentlich am stärksten betroffen. Das heißt, was jetzt so ein prädiktives System macht, ist mehr oder minder anhand von gelernten Trainingsdaten eine Vorhersage zu treffen anhand eines neuen Bildes, um zu sagen, okay, diese Pixel hier würde ich jetzt als problematisch klassifizieren und deswegen ist sie ja auch schön rot markiert. Und das gibt es in allen möglichen Facetten, in allen möglichen Erkrankungen und auch in der Radiologie in allen möglichen Bereichen von radiologischer Befundung. Das gibt es für Knochenbrüche, das gibt es jetzt hier für Stroke, das gibt es für Lunge, das gibt es für alles Mögliche, sagen wir es mal so. Und was jetzt meine Forschung auszeichnet, ist die Fragestellung, okay, wir haben jetzt solche Systeme, die machen jetzt hier zum Beispiel so eine Heatmap. Was macht das denn eigentlich mit den Radiologen, Radiologinnen, die solche Bilder sehen? Wie beeinflusst das denn eigentlich deren Entscheidungen? Solche Systeme gibt es, wie gesagt, sehr, sehr viele davon sind als Medizinprodukt zugelassen und es gibt alle möglichen Studien dazu. Es gibt eine Diskussion darüber, wie sehr werden die tatsächlich im Klinikalltag eingesetzt, da ist es ja so ein bisschen gespalten. Die meisten

Systeme, die es gibt, sind natürlich in solchen klinischen Trials, also in Versuchen, kontrollierten Einsatzes, wo die dann eingesetzt werden und dann evaluiert werden innerhalb der klinischen Praxis. Aber ja, es gibt sie. Und es gibt sie schon länger. Am Beginn einer Doktorandenzeit, also 2017, ist es mit einem Zitat mehr oder minder begonnen, das mich bis heute begleitet, von Jeffrey Hinton, der gesagt hat, okay, die Berufsgruppe der Radiologen ist wie ein Kojote, der über diesen Rand einer Klippe schon drüber ist und eigentlich keinen Boden mehr unter den Füßen hat. Weil die Algorithmen, also sein Zitat ist vor zehn Jahren her, jetzt mittlerweile in fünf Jahren so gut sein werden, dass wir die Radiologen, Radiologinnen nicht mehr benötigen. Zehn Jahre später, ja, wir haben immer noch Radiologinnen, da gibt es glaube ich auch noch so ein Meme, das ist irgendwie so ein Bild von Radiologinnen, die im Stau stehen, glaube ich, um irgendwie zur Arbeit zu kommen. Also wir sehen, dass die Berufsgruppe immer noch da ist. Wir sehen, die Algorithmen sind gut geworden, aber noch nicht so, dass sie jetzt flächendeckend überall eingesetzt werden können und flächendeckend alles machen können. Was wir aber auch sehen, ist halt der letzte Satz, dass hier viel passiert, also dass auch junge Ärztinnen und Ärzte oft die Entscheidung nicht treffen, okay, sie wollen in die Radiologie, weil sie natürlich auch Befürchtungen haben, okay, vielleicht jetzt nicht in fünf Jahren, vielleicht nicht in zehn Jahren, aber vielleicht in 20 Jahren haben wir Systeme, die das so gut können, dass ich jetzt diese Ausbildung nicht machen werde. Gleichzeitig sind wir natürlich im medizinischen Diagnostikkontext, auch in der Radiologie, im sogenannten Hochrisiko-KI-Umfeld. Das bedeutet, dass wir laut EUAI-Acts einen sogenannten Human Oversight sicherstellen müssen. Das heißt, solche Systeme dürfen keine Entscheidungen treffen, also die Systeme müssen ihre Entscheidung wie eine Art Vorschlag formulieren oder so eine Art Hilfestellung sein. Und die letztliche Verantwortung für die Diagnostikentscheidung muss von einem Menschen getroffen werden, laut diesem Oversight-Paragrafen. Das bedeutet, dass am Ende immer so ein Human, wir nennen das Human in the loop, irgendwie mit dabei sein muss, der dann am Ende dann sagt, okay, ich stimme dem zu, was das KI-System da vorschlägt, oder ich würde dem widersprechen, ich denke, da ist was anderes. Also da passiert ganz, ganz

viel. Jetzt kommt der Teil, den ich versprochen habe, mit der Forschung. Wir wissen allerdings, dass diese Aufgabe, also dieses Überprüfen von dem KI-System, das Abwägen von dem KI-System super schwierig ist für Menschen. Die können das nicht so gut. Und zwar gibt es eine vor zwei Jahren veröffentlichte Meta-Analyse, die hat sich ganz, ganz viele verschiedene Experimente angeschaut, die verschiedene Oversight-Themen hatten. Und was diese Meta-Analyse gezeigt hat, ist, ja, Mensch mit KI zusammen ist besser als Mensch allein. Vor allem, wenn der Mensch alleine nicht so gut ist, ja, dann sehen wir hier, ganz, ganz viele Studien zeigen, die Kombination ist besser. Aber Mensch mit KI zusammen ist nicht unbedingt besser, als wenn das KI-System alleine die ganzen Diagnostik oder die ganzen Entscheidungen treffen würde. Nicht so gut. Und hier ist das genau das Spannungsfeld, weil das ist ja natürlich, hier sprechen wir von Oversight, hier sprechen wir von Komplementaritäten zwischen Menschen und KI-System, ja. Schafft es der Mensch irgendwie Fehler richtig zu erkennen oder richtige Vorschläge richtig zu erkennen und gleichzeitig seine komplementären Stärken einzubringen oder nicht? Und hier sehen wir, oft ist das schwierig und oft klappt das nicht. Und das ist tatsächlich so das, was so meine Forschung in dem Bereich auch tatsächlich motiviert, ist zu verstehen, ja, warum nicht? Was passiert da? (...) Wir haben schon sehr, sehr früh damit angefangen. Also das ist jetzt eine Studie, die schon, ja, würde man sagen, jetzt veraltet, 2021 publiziert ist, jetzt schon natürlich im KI-Zeitalter alt. Aber da haben wir tatsächlich gesehen, dass es ja mehr oder minder so eine Art Prozess gibt, wenn die Leute so durchlaufen, um so Systeme zu evaluieren. Das haben wir im medizinischen Kontext getestet. Und das ist so ein Muster, was man so eine Art Deliberate Consideration, haben wir das getauft, ja, also das ist irgendwie sehr, sehr komplex. Also da muss eine eigene Entscheidung gebildet werden, erst mal ein eigenes Assessment, dann schaue ich mir dieses KI-System, diesen KI-Vorschlag an. Wenn da jetzt zum Beispiel irgendwie ein Widerspruch ist, fange ich an, mich selber ein bisschen zu reflektieren und zu hinterfragen, frage auch, okay, wie ist eigentlich die KI an dieses System, ja, wie ist es denn eigentlich zu dieser Einschätzung gekommen? Vielleicht liegt es an den Trainingdaten, vielleicht ist das Bild falsch rum drin, vielleicht ist es ein seltener Fall oder

vielleicht ist es irgendwie was anderes, was ich über das KI-System weiß. Und am Ende, basierend auf diesem sehr, sehr komplexen Hin und Her zwischen meiner Selbstreflektion und dem Hinterfragen der KI, treffe ich am Ende eine finale Entscheidung. Klingt toll, ist aber sehr, sehr anstrengend. Also wenn man sich vorstellen würde, jede einzelne Entscheidung würde all diese Schritte laufen, dann wären wir einfach, keine Ahnung, mit jedem Fall irgendwie fünf Minuten beschäftigt. Vor allem das hier, das sind sogenannte, das werde ich Ihnen gleich nochmal zeigen, metakognitive Prozesse, die brauchen einfach ganz, ganz viele kognitive Ressourcen. Also das ist, sagen wir es mal so, das, was wir idealtypisch irgendwie sehen wollen, wenn wir Oversight sehen wollen, ist aber natürlich einfach super anstrengend und keiner macht das. Stattdessen, was wir oft sehen könnten, ist der sogenannte Automation Bias-Effekt. Ja, also wir machen eigentlich nichts selber, schauen das System an und sagen, ja passt schon. So, finale Entscheidung, Abkürzung gelungen, keine Zwischenschritte notwendig, wenn das System richtig ist, haben wir eine richtige Diagnoseentscheidung am Ende, wenn das System aber falsch lag, ja, dann haben wir halt leider einen Fehler gemacht. Das ist dieses Automation Bias-Phänomen, das tatsächlich im Extremfall bedeutet, dass Leute gar nicht mehr selber denken und gar nicht mehr selber ein Urteil fällen. Das gibt es in ganz vielen verschiedenen Domänen mit Systemen, die sehr gut sind, sehr lange zuverlässig sehr gut funktionieren, weil dann die Menschen einfach aus Ressourcengründen Aufmerksamkeit verlieren und einfach nicht mehr in der Lage sind, so viel sich damit zu involvieren. Das ist ein ganz normales, sagen wir es mal so, Phänomen aus ganz vielen verschiedenen Bereichen. Es gibt es aber auch so, dass ich zum Beispiel feststelle, ah KI-System passt nicht, das liegt an dem System, das macht immer halt Fehler. Dann habe ich auch eine finale Entscheidung, ich lehne einfach das System in Anführungsstrichen ab, das ist auch eine tolle Abkürzung, ich muss mich nicht mehr selber hinterfragen, muss nicht mehr reflektieren, bin einfach schneller im Prozess. Und ich weiß, das System funktioniert nicht, ich weiß, es hat Fehler. Super gutes Muster, wenn wir wissen, dass das System sehr fehlerhaft ist, dann kann ich es einfach wegnorieren, muss ich nicht wirklich im Detail nochmal drüber gehen, drüber

nachdenken, hilft, bin ich schnell, effizient. Problematisch, wenn das System gut ist und recht hat und dummerweise das jetzt ein Fall ist, wo tatsächlich das System einen richtigen Vorschlag hatte und ich vielleicht irgendwas übersehen habe. Diese Fälle gibt es, das sind die sogenannten Underreliance-Fälle. (...) So, wir haben ja so gesehen, das war ja 2021, da haben wir jetzt ganz viel gesehen, dass solche Entscheidungsprozesse komplex laufen, aber natürlich die Leute halt auch gerne mal wieder eine Abkürzung wählen. Was wir sehen oft ist, um die Leute zu unterstützen, das zu tun, machen viele Leute, sagen wir es mal so, Erklärungen. Ja, es gibt verschiedene Arten und Weisen, KI-Systeme zu erklären, das ist jetzt ein Scherbdigramm, also das es einfach verschiedene Werte zeigt, es gibt generative KI-Systeme, die Erklärungen generieren und dann sagen, okay, warum würde ich jetzt einen bestimmten Vorschlag machen und es gibt natürlich auch verschiedene Heatmap zum Beispiel. Das ist wichtig, weil das den Leuten halt erlaubt, vielleicht auch mehr diese Reflexionsfähigkeiten zu haben, aber gleichzeitig wissen wir nicht so genau, ob das denn eigentlich noch mehr zu diesem Overload beiträgt, den ich jetzt beschrieben habe und noch mehr zu Abkürzungen führt. So, jetzt habe ich versprochen, ich gucke nur auf die Zeit, dass ich nicht zu lange rede, aber ich habe versprochen, ich bringe ein bisschen Psychologie mit und ich bringe ein paar Fachbegriffe mit, die wirklich in das Tiefen des menschlichen Denkens gehen. Und das ist eigentlich hier meine Lieblings-Slide, um Ihnen das mitzubringen. Wenn wir vom menschlichen Denken besprechen, ich hatte das Wort schon genannt, das ist das sogenannte Wort der Metakognition. Was ist das eigentlich? Also man kann sich so vorstellen, das ist das Denken über das Denken. Es läuft nicht immer. Also normalerweise, wenn ich mit einer Aufgabe beschäftigt bin, denke ich halt über diese Aufgabe nach, versuche diese Aufgabe zu lösen, versuche dann verschiedene, wir haben vorhin gehört, klinische Algorithmen durchzugehen, versuche verschiedene Indikationen mir anzunehmen und bin einfach in der Aufgabe. Also da habe ich auch meine Heuristiken, aber ich bin eigentlich in der Aufgabe mit dabei. Was die Metakognition macht, ist im Moment zu dem normalen Prozess zu stoppen. Also wenn ich jetzt zum Beispiel meine Aufgabe mache und feststelle, irgendwas stimmt nicht. Ich mache eine

Pause. Ich merke, ich werde unsicher. Ich merke, irgendwo ist ein Fehler. Ich fange an, nochmal zu hinterfragen, habe ich denn jetzt gerade richtig gemacht? Habe ich denn jetzt vielleicht irgendwas übersehen? Was passt denn jetzt hier eigentlich zusammen? Oft ist das so ein diffuses Gefühl und das ist so ein metakognitiver Eingriff in meinen normalen Prozess irgendwo. Das messen wir oft mit meiner Sicherheit, also eine Art, wie sicher bin ich mir jetzt eigentlich? Ich werde auf einmal unsicher und das ist oft so ein Indikator von, okay, da passiert was auf einer Metaebene. Ich beobachte mich jetzt gerade selber im Denken und ich versuche, mein eigenes Denken zu steuern. Wie mache ich das? Ich kann zum Beispiel, keine Ahnung, ich merke, ich bin unsicher. Ich könnte einen Kollegen, Kollegin fragen. Ich könnte nochmal das KI-System hinzuziehen. Ich könnte aber auch nochmal jemanden eine Zweitmeinung bitten, ein Team bitten und so weiter. Das sind oft so Kontrollstrategien, nochmal alles durchgehen zum Beispiel. Das sind solche, sagen wir es mal so, metakognitive Kontrollstrategien. (...) Für den weiteren, diese metakognitiven Kontrollstrategien können gut sein, können auch nicht so gut sein. Was wir uns anschauen, ist die sogenannte, wenn wir eine Übereinstimmung haben zwischen, wie gut kann ich mich eigentlich selber beobachten und wie gut bin ich denn tatsächlich in der Entscheidung. Das heißt, wir messen das oft, indem wir Confidence und Accuracy in einen Zusammenhang setzen. Zum Beispiel, ich bin mir sehr sicher in einer bestimmten Entscheidung, zum Beispiel in einer Diagnostikentscheidung, dass die jetzt richtig ist. Ich bin auch richtig. Das Ergebnis passt. Das heißt, hier habe ich eine hohe Übereinstimmung auf einer metakognitiven Ebene, weil ich bin richtig, ich bin mir sicher, meine Sicherheit ist hoch, mein Ergebnis passt. Wenn ich jetzt aber zum Beispiel mir sehr sicher bin, aber völlig falsch liege, das Phänomen kennen wir alle, das ist dieser Overconfidence-Effekt. Ich bin mir sehr sicher, aber bin komplett daneben. Keine so gute Übereinstimmung zwischen Confidence und Accuracy. Das messen wir anhand von der Kalibrierung und der Resolution entweder auf einer Fall-zu-Fall-Basis auf jeder bestimmten Entscheidung oder über den gesamten Kontext hinweg. (...) So, wir wissen, diese Metakognitionen lassen sich sehr, sehr schnell beeinflussen. Ich überspringe das mal, aber wir wollten eigentlich auch wissen, ja, okay,

jetzt habe ich diese Metakognition. Was macht das denn eigentlich, wenn ich mit einem KI-System interagiere? Und das ist eigentlich so eine Studie, die ich jetzt mitgebracht habe, dass es so zwei Studien sind, zwei Experimente, so ein bisschen, um auch zu illustrieren, wie wir eigentlich immer diese Kollaboration oder diese Oversight-Challenge, von der ich am Anfang gesprochen habe, eigentlich messen. Was wir dafür brauchen, ist, wir müssen wissen, ob eine Diagnose richtig ist oder falsch ist. Weil sonst können wir nicht diese ganzen Kallibrierungssachen messen, sonst können wir kein Oversight messen. Das heißt, was wir machen, wir erzeugen eine künstliche Umgebung, eine sogenannte Experimentalumgebung, in der wir relativ genau festlegen, wann sind wir, also wann ist denn zum Beispiel das KI-System richtig, wann ist das KI-System falsch, ist das ein schwieriger Fall, ist das ein einfacher Fall und was ist eigentlich die Enddiagnose, die hier in diesem Fall vorliegt, ja? Das machen wir sehr, sehr kontrolliert oft, also hier hinten dran sind wir drei Radiologen, die sich dann hingesetzt haben und die Bilder in einem Konsens befunden haben, ja Also das ist ein sehr künstliches Umfeld, einfach nur, weil wir wissen wollen, wie gut ist die Entscheidung und das können wir natürlich im Klinikalltag nicht machen, weil oft einfach auch nicht klar ist, was eigentlich am Ende richtig ist, ja? Hier können wir es kontrollieren und in einem Klinikalltag können wir auch nicht falsche KI-Vorschläge reinschmuggeln, um zu gucken, welchen Einfluss hat das, das wäre absolut unethisch, ja? Hier können wir das. Also hier haben wir falsche KI-Vorschläge, hier haben wir diese Erklärungen, die ich ja vorher erwähnt habe, wir wollten wissen, was machen Erklärungen auch auf dieser Metakognitionsebene, was macht dieser Vorschlag und was machen die Leute eigentlich einfach so? Das ist ein sehr, sehr, sehr großes Experiment, was gelaufen ist, was jetzt endlich akzeptiert wurde, was tatsächlich 150 Entscheidungen pro Person hatte. Wir hatten 44 Experten, 113 Novizen, also können Sie sich vorstellen, wie viele Entscheidungen wir da insgesamt hatten, dass Sie sich alle mal zusammen multiplizieren. Was wir hatten in diesem künstlichen Setup war die Diagnose von Hirnblutungen. Und unsere Erklärungen waren so Bounding-Boxes, die in einer Schicht von einem CT-Bild des Gehirns drin waren, um zu zeigen, wo denn da eine Blutung ist. Wie gesagt, sehr künstlich, mit Absicht künstlich,

um wirklich auch festzustellen, was für einen Einfluss hat das denn auf die Diagnostikqualität. Und dann sollten die Leute bewerten, wie sicher sind sie und das brauchen wir, wie ich vorher erklärt habe, für unsere Metakognitionsgeschichten. So, was sehen wir? Relativ typisch, relativ, sagen wir es mal so, nicht weltbewegend an dieser Stelle. Wir sehen, Novizen, also Leute, die wenig Expertise haben, werden besser mit KI, werden viel, viel besser, wenn sie noch eine Erklärung dazu bekommen. Experten profitieren weniger von einem KI-Vorschlag in unserem Setup, wenn dann nur das mit der Bounding-Box, aber auch nicht so stark wie bei den Novizen. Hier ist auch nochmal ein sehr, sehr typisches Ergebnisbild, was wir in ganz vielen Studien sehen. Sie sehen das hier, was passiert eigentlich, wenn KI-Systemvorschlag richtig oder falsch ist. Wir haben das ja manipuliert experimentell. Da sehen wir, bei den Experten löste auch ein falscher Vorschlag einen Drop aus in der Genauigkeit. Sie lassen sich davon, die sind nicht so gut darin, das zu identifizieren. Ein positiver Vorschlag erhöht die Genauigkeit auch statistisch signifikant, aber leider nicht so sehr, wie wir das haben wollen. Bei den Novizen ist es ja all over the place. Das springt sehr, sehr stark. Also das ist ein relativ normales Bild, das findet man ganz, ganz vielen Studien, dass Novizen sich extrem stark von solchen Systemvorschlägen leiten lassen. Also sehr starker Zugewinn in der Accuracy, wenn das System richtig ist, und sehr, sehr, sehr starker, extrem starker Abfall, wenn das System einen Fehler macht. Ganz typisch. Die Novizen gewinnen auch mehr an diagnostischer Sicherheit, sind sich aber natürlich insgesamt auch weniger sicher. So, was macht das denn jetzt eigentlich mit diesen Metakognitionen, von denen ich ja so erzählt habe, wie wichtig die sind? Also wir haben ein standardisiertes Maß dieser Metakognition genommen und haben mal geschaut, okay, Leute, die sich besser in dieser Metakognition kalibriert sind, also Leute, die sich ohne KI besser einschätzen können, sind das auch die Leute, die vielleicht hier besser mit so einem KI-System interagieren können? Was wir sehen ist, Experten können sich generell besser einschätzen als Novizen. Höhere Werte bedeutet bessere Kalibrierung in dem Sinne oder Resolution in dem Fall. Novizen sind ein bisschen mehr verteilt, sie können sich selber noch nicht so gut einschätzen, das ist jetzt auch noch nichts Neues. Aber was wir

sehen, ist, dass Experten, die in der Lage sind, sich besser selber einzuschätzen, auch diejenigen sind, die am besten von dem KI-System profitieren. Das heißt, wir sehen einen Interaktionseffekt zwischen einem Fähigkeitseffekt und diesem metakognitiven Effekt. Das heißt, wir brauchen Leute, die in der Lage sind, unabhängige Entscheidungen zu treffen und die noch zusätzlich dazu in der Lage sind, zu reflektieren, um dann wirklich das meiste aus so einem System herauszuholen. Ich habe jetzt einfach mal ein paar Regressionen mitgebracht. Entschuldigung dafür, aber was ganz spannend ist, diese diagnostische Sicherheit, also die Diagnostic Accuracy ohne KI, hat einen großen Effekt darauf, wie gut ich mit KI arbeite. Unabhängig davon, ob ich Experte oder Novize bin. Und die Erklärungen helfen den Novizen mehr, während die bessere Selbstreflexion den Experten mehr hilft. Also da sehen wir so einen differenzierten Effekt hier, dass die Explanations von dem KI-System doch eher mehr den Novizen helfen, vielleicht auch zu lernen oder zu finden, wo denn irgendwas komisch ist. Aber die Selbstreflexion viel wichtiger ist für die Experten. (...) So und was macht diese Metakognition? Die hilft vor allem, sagen wir es mal so, eher ein bisschen deskriptiv, explorativ, weil unsere Daten dafür nicht ausreichen, leider eher bei den falschen KI-Hinweisen und ein bisschen weniger bei den korrekten Hinweisen. So, ich will jetzt nicht zu viel reden, aber ich glaube zwei Sachen, die ich noch sagen möchte. Das, was wir jetzt gemacht haben, war ein sehr, sehr standardisiertes Experiment. Und wir haben angenommen, dass Menschen und KI irgendwie ähnlich Entscheidungen treffen. Aber wir wissen, dass das eigentlich gar nicht der Fall ist. Also was ein KI-System macht, hatte ich ja gesagt, macht out of sample Vorhersagen anhand von Datenpunkten und Pixeln. Was Menschen machen, ist Kontext sehen, Patientenfälle sehen. Zum Beispiel hier ist eine Blutung, weil ein ähnliches Muster bei Blutungen in den Trainingsdaten ist. Das wäre, was die KI machen würde. Was jetzt aber ein Mensch machen würde, es wäre ein älterer Patient hier. Ich sehe ganz klar, da ist was, er sofort innerhalb von Sekunden, wenn es ein Experte ist. Und die Ursache davon, ach, die Person muss wohl gestürzt sein, ja, all das so denken KI-Systemen nicht. Den KI-Systemen denkt er, da ist irgendwas komisch, deswegen sind da ein paar Pixel ein bisschen heller, klassifiziere ich jetzt als Blutung.

Und das ist etwas, was wir natürlich in unseren Experimenten nicht sehen, dieses einfach, dieses kontextuelle Denken, was Experten, Expertinnen mitbringen. Und natürlich, was wir jetzt gemacht haben, ist eine ganz, ganz starke Abstraktion. In Realität ist natürlich der Alltag komplett anders. Also das ist jetzt so ein Foto von so einem typischen Workplace von der Radiologin, Radiologin, da passiert einfach ganz viel drum herum. Da kommt ein Notfall, da klingelt ein Telefon, die Arbeit wird unterbrochen, dann wird auf einmal ein neuer Fall geöffnet, dann kommt noch mal jemand rein, will nochmal Hilfe haben. Parallel befundet man, während man diese Aufgabe macht. Das heißt, in Realität ist natürlich die Arbeit wesentlich komplexer, als jetzt einfach nur ein bestimmtes Bild zu befunden, was wir da jetzt angeschaut haben. Zu sagen, ist da eine Blutung? Ja, nein, das macht keiner. Für uns hilft das natürlich experimentell ganz viel zu abstrahieren, aber hier ist natürlich auch einfach wesentlich mehr Komplexität dahinter. (...) Einfach nur noch ein kleiner Ausblick, ich glaube, ich habe noch ein paar Minuten. Ich hatte jetzt ganz viel von diesem sogenannten prädiktiven System gesprochen, was wir natürlich sehen ist, was passiert jetzt mit dieser Blackbox-Generative-AI? Und da sehen wir sehr, sehr viele tatsächlich, dass solche Systeme in die medizinische Realität kommen und Aufgaben machen, die wir vorher diese prädiktiven Systeme hatten. Also hier ist zum Beispiel, ja, okay, hier wird einfach beschrieben, was auf dem Bild ist. Da wird automatisiert ein Befund erzeugt. Gleichzeitig auf Patientinnen und Patienten Seite wird einfach sehr, sehr viele Daten, medizinische Daten, spezifische KI-Systeme, immer tatsächlich darauf trainiert, nochmal hier mit Patienten und Patientinnen zu kommunizieren. Wichtig hier, tatsächlich sind medizinische Fälle der größte Use Case. Also Therapy, ja, ich habe das in meiner Medizin mal zusammengefasst, Companionship-Cases, ja, ist Top 1 Use Case von ChatGPT, von Menschen. Also das muss man sich, also solche Sachen, wo die Leute wirklich persönliche Informationen teilen, auch medizinische Fragestellungen diskutieren, ist das tatsächlich die Realität von Patienten und Patientinnen, die dann tatsächlich danach in die medizinische Praxis kommen. Warum sind die ganzen Ergebnisse von diesen Experimenten, die ich vorher gezeigt habe, hier trotzdem wichtig? Weil natürlich solche Systeme auch Fehler machen.

Sie halluzinieren, deren Fehler sind noch komplexer zu identifizieren. Und langfristig, also auch mit diesem Zitat von Jeffrey Hinton, das ist natürlich die Frage, okay, wie bringen wir das eigentlich den Leuten bei, diese Fehler zu identifizieren? Wie bringen wir denen eigentlich genug medizinische Fähigkeiten bei, genug Selbstreflexion bei und erlauben sie gleichzeitig tatsächlich zu lernen von den Systemen, damit sie vielleicht auch besser werden? Also es ist ein relativ komplexes Feld, was wir hier haben. (...)

Genau, Fazit, Metakognition, also KI-Unterstützung kann diagnostische Aufwand reduzieren, aber wir haben ja gesehen, das ist super aufwendig, dieses ständige Hin und Her, dieses Selbstreflektieren, das macht keiner den ganzen Tag, einfach anstrengend kognitiv. Deswegen ist es klar, dass wir irgendwann einfach sogenannte Biases haben und einfach irgendwie vielleicht auch mehr der Automatisierung folgen. Die Frage ist, kommt man da rechtzeitig wieder raus? Das ist immer so die Frage, die aus einer metakognitiven Perspektive immer wichtig ist. Menschlicher Oversight ist halt sehr, sehr herausfordernd und schwierig und auch oft so nicht unetzbar aus einer Kognitionsperspektive. Da muss man auch überlegen, wie kann man das regulatorisch vielleicht umgestalten, dass nicht die ganze Verantwortung einfach auf Ärztinnen und Ärztinnen lastet, diesen Oversight dann irgendwie sicherzustellen, weil das führt dazu, dass die Leute natürlich das in der Praxis nicht nutzen. Also sie müssen nicht nur selber befunden die ganze Zeit, weil sie müssen ja immer das KI-System ja noch hinterfragen, sondern sie müssen auch noch die Verantwortung übernehmen, falls dieses System Mist baut. Also es setzt halt einfach sehr, sehr viele negative Anreize einfach für Leute in der Praxis tatsächlich. Und gleichzeitig haben wir einen Bereich, der absolut nicht reguliert ist, wie ich zum Beispiel zu Hause meinen medizinischen Befund in ChatGPT reingebe, um mir daraus dann zu sagen, was mache ich denn als nächstes? Absolut nicht regulierter Bereich, wird sehr, sehr stark genutzt. Das heißt, diese Oversight-Thematik aus regulatorischer Perspektive muss man tatsächlich nochmal kritisch durchdenken, wie man das da jetzt rangeht. Und Forschungsprogramm, wie ich gesagt habe, unsere Experimente sind eigentlich ganz toll, aber immer eine Abstraktion der Wirklichkeit. Die Realität ist komplexer, die Entscheidungen sind komplexer, es passiert ganz viel

drumherum. Das heißt, wir brauchen mehr Studien tatsächlich zum Einsatz von KI in der Praxis und wir brauchen, ich habe es jetzt mal „shadow GPT“ genannt, Berücksichtigung von generativer KI. Selbst viele Leute nutzen es auch, als Medizinerin habe ich ganz viele Beispiele gesehen, obwohl das nicht offiziell zugelassen wird. Viele, viele Leute nutzen das einfach parallel on top. Was macht das denn eigentlich mit uns? Was macht das denn eigentlich in der Diagnostik? Im besten Fall führt es dazu, dass man besser wird, aber wir wissen, dass es immer natürlich auch noch eine zweite Seite geben kann. Und das heißt, wir brauchen mehr so eine Komplementarität zwischen Feldstudien und Experimenten. In dem Sinne bedanke ich mich ganz herzlich für die Aufmerksamkeit und freue mich auf Diskussionen und Fragen. Ganz herzlichen Dank. (...) Ja, prima. Tolle Keynote auch. Kommen wir zu Fragen. Rückfragen, Gesprächen. Hallo, hi, Justin Fishedick von der Uni Göttingen. Danke für den spannenden Vortrag. Vielleicht um direkt mal so ein bisschen die Brücke zu dem vorherigen Vortrag zu schlagen, wo es ja so ein bisschen um die Ausbildung von angehenden Medizinerinnen und Medizinern ging. Ich weiß aus meinem nächsten Umfeld, dass so was wie KI-Systeme es ins Curriculum von so einem Medizinstudium, zumindest in dem Fall jetzt noch nicht so geschafft haben. Jetzt meinten Sie ja in Ihrem Vortrag unter anderem, diese Biases, die kommen ja sowieso. Das kann man nicht verhindern. Gibt es da vielleicht trotzdem quasi eine Möglichkeit, zum Beispiel innerhalb der Ausbildung von angehenden Medizinerinnen und Medizinerinnen, die so zu schulen, dass vielleicht der Einfluss dieser Biases nicht so stark ist? Dass man vielleicht die Over-Reliance oder Under-Reliance, vielleicht auch im Sinne von einer AI-Aversion oder Appreciation, dass man das ein Stück weit ausgleicht? Wo sehen Sie da vielleicht die Möglichkeiten, da auch ein Stück weit früher anzusetzen, um präventiv einen besseren Umgang herzustellen? Ja, sehr wichtiger Punkt. Das habe ich auch schon sehr, sehr häufig gehört. Ich weiß, dass hin und wieder Kolleginnen und Kollegen das einstreuen. Ich glaube, das Wichtigste ist tatsächlich auch, ich vermute ähnlich wie in der Chirurgie, dass man den Umgang üben muss. Dass man testen muss, okay, wie reagiere ich denn eigentlich? Was kann das System, was kann das nicht? Und auch ein bisschen dieses Demystifizieren von KI. Viele denken, dass es so ein, ja, wie so,

keine Ahnung, so Gottin Weiß in Anführungsstrichen, in der Lage, irgendwie alles zu tun. Es gibt sehr, sehr viele KI-Algorithmen, die man gar nicht mitbekommt. Also wenn ich jetzt zum Beispiel einen CT-Scan mache und ein modernes Gerät habe und das Bild wird rekonstruiert, im Hintergrund wird einfach nur die Schärfe von dem Bild erhöht. Ja, das sind auch KI-Systeme im Hintergrund. Es sind also viele Systeme, die man gar nicht so mitbekommt. Das heißt, da einfach nochmal zu sagen, okay, was gibt es denn konkret? Was können die Systeme, was können sie nicht? Wie stellen wir sicher, das ist eine Frage an die Mediziner, Medizinerinnen im Raum, würde ich mal sagen, ja, wie stellen wir eigentlich sicher, dass genug medizinische Fähigkeiten vorhanden sind, um diese Oversight überhaupt durchzuführen? Weil das ist, glaube ich, so ein Hauptergebnis, was wir in dieser Studie haben, ist einfach, dass die so wichtig sind. Ja, wenn ich ohne KI gut bin, bin ich auch mit KI besser. Aber trotzdem, diese Basis muss da sein. Wenn die Basis nicht da ist, dann laufe ich natürlich in diesen Automation Bias rein. Also ich bin ein völliger, 100-prozentiger Delegator von KI. Wenn ich ein Medizinbefund habe, ja, natürlich schicke ich das an, in Anführungsstrichen, natürlich schicke ich das an Chat GPT, aber unabhängig, was Chat GPT mir sagt, ich kann das nicht verifizieren. Ja, ich denke mir so, ja, ja, das klingt plausibel. Woher soll ich es wissen? Ich habe nicht die Ausbildung, ja. Und hier einfach in diesen Dialog zu treten, okay, also es gibt einfach, ja, sehr, sehr viel. Und im Endeffekt ist es das, was Patienten und Patientinnen tun, ja. Die kommen einfach mit diesem Wissen. Die kommen mit mehr Wissen. Und die kommen tatsächlich, die Modelle sind so gut, dass die mit sehr gutem Wissen zum Teil kommen. Das heißt, das ist auch eine Sache, die muss man den Leuten beibringen auch, ja. Aber da bin ich leider an der falschen Stelle, um das zu tun, ja. (...) Vielen Dank für den tollen Vortrag. Ich glaube, wir müssen uns mal vernetzen, dass wir da analysieren, zusammenarbeiten können. Ein Punkt, der uns auch gerade schon beschäftigt, ist unsere Lernenden, die angehenden Mediziner sind ja nicht Blanco. Im Moment ist ja der Idee oder so ein bisschen so die Formulierung, ja, dann kommen die rein und irgendwann haben die KI. Aber die lernen ja jetzt schon nur noch mit KI. Die nehmen keine Bücher mehr, die nehmen kein E-Learning, sondern die quizzten sich mit KI. Wir machen da

gerade ein paar Untersuchungen zu und schauen, wie die lernen, was die machen. Und die nutzen ja konstant KI. Das heißt, diesen Punkt, ich habe eine gewisse Expertise, damit ich beurteilen kann, ob das, was die KI mir sagt, den haben wir nicht mehr. Und als schwarzes Bild zum Zeichnen, und dann müssen wir mal gucken, wie wir da weiterkommen. Wir waren die Ersten, die die Covid-Patienten hatten in Frankfurt und wir haben eine sehr hohe Expertise in der Mikrobiologie. Und ich weiß, ich muss den anrufen und sage, ich habe einen Patienten mit dem, dem und dem. Dann sagt er, okay, das Antibiotikum, dies, egal, welche Resistenzen sind, was. Das ist ein Experte, weil er über 20 Jahre Expertentum hatte. Jetzt nutzt er KI und hinterfragt die aber auch kritisch. Aber unsere Junioren, die kriegen diese Expertenausbildung nicht mehr, weil sie konstant schon mit KI lernen und lernen nicht mehr, diese klinischen Befunde, die sie ja auch gesagt haben, zu kriegen. Und da ist ein Punkt, an dem wir schauen müssen, wie kriegen wir noch eigenes Expertentum, damit wir überhaupt KI in drei, vier, fünf Jahren in der Medizin noch bewerten können. Da hoffe ich, dass Sie da auch irgendwie ein paar Antworten drauf finden. Ja, es klappt so rein. Von der Forschung gibt es, glaube ich, zwei Theorien. Es gibt die eine Theorie dieses Hyperskillings, die werden schneller, besser. Und dann gibt es die zweite Theorie des Neverskillings, in Anführungsstrichen, die lernen es nie. Und irgendwie, es gibt sehr wenig empirische Studien. Also wir wissen es nicht. Wir wissen, wir wissen es einfach nicht, was die machen müssen. Das betrifft ja nicht nur die Medizin, das betrifft ja auch Schulen, das betrifft ja auch Schüler, Schülerinnen. Wenn man das aus metakognitiver Perspektive betrachtet, das kriegt man nur, wenn man es ohne macht. Also diese metakognitiven Reflexionsfähigkeiten bekommt man nur ohne KI. Und das muss man auch mal überlegen. So, okay, was heißt das denn eigentlich? Wo ist es denn okay, dass Sie mit KI lernen und wo müssen Sie es ohne machen? Also das ist ein sehr, sehr komplexer Aushandlungsprozess. Ja, das wäre auch so ein bisschen meine Frage gewesen. Sie haben ganz gut gezeigt, dass die Novizen gegenüber Novizen mit KI besser werden als, dann hatten Sie die Experten. Die Frage ist, gibt es Arbeiten, wo man sieht, ob aus Novizen mit KI die Experten werden, die Sie jetzt verglichen haben, die noch ohne KI aufgewachsen sind? Weil ich glaube, das wird uns in

Zukunft beschäftigen. Wie lernen in Zukunft die Generationen und wie wird man zum Experten? Das gibt es sehr, also nichts, glaube ich. Gibt es nichts, ne? Also nicht, dass ich was wüsste. Also es gibt, wie gesagt, es ist eine Diskussion, die wir jetzt gerade in ganz vielen Bereichen auch führen, mit ChatGPT und lernen, ob wir das können. Und es ist komplex. Die zweite Frage ist, weil Sie sagten, so der Haupt-Use-Case ist, dass Patienten oder medizinische Laien sehr viel ChatGPT oder andere Systeme befragen. Gibt es da Untersuchungen zu, was das mit der Lernfähigkeit macht? Also wie verschiebt sich sozusagen das Lernen der KI durch diese zahlenmäßig vielen Konversationen gegenüber dem Expertenwissen? Also ändert sich das? Also wir wissen, also ich meine ChatGPT hat nicht umsonst einen ganz konkreten KI-Agenten mehr oder minder im Gesundheitssystem rausgebracht. Ja, also das ist ja bei den generativen KI-Systemen so, wenn man jetzt zum Beispiel ChatGPT nutzt und erlaubt, dass die Daten genutzt werden. Ich weiß nicht, ob wahrscheinlich auch, wenn man es nicht erlaubt. Aber unabhängig davon, das Modell lernt ja anhand von den Inputs, die wir geben. Und je mehr Konversationen zu einem bestimmten Thema geführt werden, desto besser wird das Modell in diesem bestimmten Thema. Dementsprechend, wenn sehr viele Leute das halt eben für jetzt hier Companionship, Therapy oder Medical Purposes nutzen, ja, dann werden die halt besser darin. Das muss man sich halt ganz klar sein. Also deswegen, also Therapy, Companionship, das ist eine sehr, sehr große Kategorie. Ich habe da jetzt medizinische Themen mit reingenommen. Aber da betrifft ja alles, wenn die Leute einfach zum Beispiel Sorgen haben und einfach so einen Freund brauchen, ist ja auch mit drin, ja, so ein bisschen. (...) Genau, sagen wir es mal so. Ja, das ist, glaube ich, alles von ich gebe einen bestimmten Befund rein und möchte den interpretiert wissen zu ich brauche einen Freund drin. Und die Modelle werden besser in den Fällen. Und dann muss man sich langfristig auch anschauen, was das für eine Auswirkung hat. Ja, vielen Dank für den spannenden Vortrag. Ich musste ganz praktisch gerade daran denken, ChatGPT oder andere KI-Modelle machen ja so eine Art, ja, ich würde sagen Pseudokognition. Also ich denke jetzt gerade darüber nach und etc. Macht das was mit der Confidence mit den Personen? Also wenn ich sozusagen allein in ChatGPT, das

ChatGPT denkt darüber nach, jetzt verbinde ich das damit. Macht das was mit der Confidence bei Novizen und Experten? Also wenn Chat, also genau, ja, es macht auf jeden Fall was mit den Leuten. Man muss aber ganz klar differenzieren, was ChatGPT schreibt was es tut und was es tut. Also je nachdem, was für ein System man hat, wie man es konfiguriert hat, je nachdem welches Modell gerade online ist, verändert sich ja in Anführungsstrichen das Modellverhalten drastisch. Also es gibt ja ganz viele Studien, die sich wirklich nur mit diesem Modellverhalten anschauen. Und wir hatten, ich glaube, vor ein paar Generationen KI-Systeme, die so eine Overconfidence geneigt haben, also die so sehr, sehr, sehr confident waren. Ich weiß jetzt nicht, ob das Problem 100% behoben ist, aber so gefühlt ist es jetzt ein bisschen weniger so überschwänglich, so sehr, sehr confident. Ja, es macht auf jeden Fall was mit den Menschen. Also wir wissen ja von den, hier, ich habe das jetzt hier am Rande einmal mitgebracht und übersprungen. Es gibt ja diese Fälle, wo Leute dann mehr oder minder in den Suizid diskutiert wurden, einfach weil das System die sehr, sehr stark spiegelt in ihren eigenen Überzeugungen. Und das ist ein sehr, sehr großes AI-Safety-Problem. Da versuchen auch ganz viele Modellhersteller dagegen natürlich auch zu arbeiten, dass die Modelle einfach so Verhaltensweise zeigen. Was mit generativer KI das Spannende ist, ist einfach nur, dass man nicht genau weiß, wieso es sich auf eine bestimmte Art und Weise verhält. Und das kann sein, wie gesagt, es kann von Modellgeneration zu Modellgeneration wechseln, dass ein bestimmtes Verhalten gezeigt wird. In einer Modellgeneration und in der nächsten irgendwie nicht mehr. Und die Systeme sind sehr, so erfolgreich, weil die einfach drauf optimiert sind, für uns einfach zu benutzen zu sein und so ein bisschen so eine menschenähnliche Konversation mit uns zu machen. Und deswegen werden sie ja natürlich auch mehr genutzt. Also ja, die sind drauf designed, dass wir sie nutzen. Also das heißt, alles was sie tun, beeinflusst uns auch natürlich irgendwo. Also ich weiß nicht, ob die Frage, das ist ein bisschen so eine generelle Antwort auf die Frage gewesen. Aber wir haben jetzt keine Studie gemacht, wie diese Overconfidence von einem KI-System auf die Medizin wirkt. Das müssten wir nochmal testen. Ich habe noch eine Frage. Vielen Dank erst mal für den Vortrag. Die Mediziner

sind ja mit Verzerrungen, kommen die ja auf die Welt. Alle Menschen. Ich auch, ja. Und es gibt ja auch vor der KI viele Verzerrungen, mit denen Mediziner sich selbst überschätzen oder eine Entscheidung treffen zu müssen und aufzuschieben und so weiter. Gibt es dazu Daten, welche Verzerrungen von der KI am besten abgegriffen werden und welche punktuellen Verzerrungen durch die KI menschlich besser ausgeglichen werden, sodass man sozusagen das Beste von beiden Welten zusammenwirft, ohne sich selber schwarze Löcher, wo beide sozusagen reintappen, zu schaffen? Ein sehr, sehr guter Punkt. Ich überlege gerade, wie ich darauf antworte. Wer wir tatsächlich so ein bisschen Blick haben, welche neuen Verzerrungen eigentlich ausgelöst werden, wenn ich ein KI-System habe. Was wir sehen, auch bei einer anderen Studie, die relativ faszinierend ist, wir sehen, dass Leute zum Beispiel sehr so ein bisschen so ein aversives Verhalten haben, nach dem Motto, ich möchte auf gar keinen Fall, gar keinen Fall von diesem System beeinflusst werden. Ich schaue auf jeden Fall 100 Prozent immer zuerst selber drauf. Also wir hatten so ein Experiment gemacht, wo die Leute wirklich das KI-System so wirklich mit der Kamera zugehalten haben, um bloß nicht davon ab, wirklich bloß nicht beeinflusst zu werden. Da sehen wir so ein bisschen so eine sehr, sehr große Angst davor, diese Verzerrungen, da, sagen wir es mal so, Opfer von diesen Verzerrungen zu werden, die neu dazukommen. Das ist eine super spannende Frage, haben wir jetzt nicht angeschaut, weil wir mehr darauf fokussiert haben, was macht dieses System und wie strukturiert sich die Kognition um dieses System herum? Ich glaube, man liest ja immer viel, dass die KI auch Fehler macht. Man lässt halt immer den Fokus raus, wie viele Fehler Mediziner eigentlich machen und nicht, weil sie so schlecht sind, sondern einfach, weil wir menschlich dazu tendieren. Ich glaube, das ist ein Punkt, den man irgendwie mehr mit reinnehmen oder es wäre schön, wenn es dazu Daten gäbe, weil das, was Sie gezeigt haben, dass der Mensch halt daran denkt, ist vielleicht ein älterer Patient und ist vielleicht gestürzt, das ist ja schon eine Verzerrung. Im Endeffekt wissen wir ja gar nicht, ob der Kontext für uns Menschen den Sinn macht in der Situation, den wir brauchen. Und vielleicht ist es ja je nach Situation besser, ohne menschlichen Kontext oder mit KI-Kontext oder mit gemeinsamen Kontext

daran zu gehen. Da sind wir wahrscheinlich auch noch relativ am Anfang zu wissen, wie man zu einer besten gemeinsamen Entscheidung kommt. Ja, also wir sehen schon, also hier, dass Experten also von diesem KI-System mit Annotation ja auch profitieren. Also offensichtlich ist diese zweite Meinung, die Sie bekommen, irgendwie, also wir haben auch ganz oft den Fall, da, ah, das habe ich jetzt nicht gesehen. Da ist jetzt noch dieses kleine Ding und dann schaut man sich das noch mal im Detail an und sagt, ja, das ist wirklich eine ganz komplexe, kleine Blutung, die hätte ich sonst übersehen. Das gibt es schon, ja. Also vor allem dieses, weil in der Radiologie ist viel dieses Sehen und Recognizen, also man muss es finden irgendwie, also wenn es klein wird, man muss es finden und man muss es richtig klassifizieren. Ja, also ist das jetzt das oder ist das das? Ich glaube, ja, also ja, man profitiert davon. Also ja, wenn es gut ist, sehen wir ganz, ganz, ganz viele Studien, die zeigen, wenn Systeme gut sind, werden die Leute besser. Also ja, deswegen ja, aber wir sehen halt immer diese Gefahr von, was passiert, wenn das System Fehler ist. Wir wissen aber auch in Realität, das System macht selten Fehler, aber es ist auch unberechenbar. Also wir haben so Phänomene wie Model Drift oder so was, dass Systeme in der Praxis auf einmal schlechter werden. Ganz normal einfach wegen den Algorithmen, die dahinter liegen. Und das ist, glaube ich, so ein bisschen, wo die Leute so ein bisschen Angst haben wegen dieser Responsibility, der halt einfach mit dazu hängt. Ja, also weil ich dann halt einfach sofort, wenn das System falsch liegt, lasse ich mich zu sehr beeinflussen, bin ich dann responsible für die Fehlentscheidung. Und das ist, ich glaube, so ein ganz, ganz komplexes Spielfeld, was man da haben, was die Leute einfach schwierig macht, so komplett von diesem System zu profitieren, wie man es könnte. (...) Ich würde gerne bei dem Punkt Responsibility noch mal einhaken, weil ich glaube, der hat auch was mit der Komplexität zu tun. Wenn wir an Krankenhäuser in der Befundung denken oder auch ein radiologisches MVZ, die Menschen, die dann die Befundung machen, das sind ja nicht immer die Chefärzte, sondern oder die Praxisinhaber, sondern das sind dann eben Assistenzärzte oder Ärzte in Weiterbildung oder junge Ärzte. Und die bekommen ein System von ihrem Chef, ihrer Chefin, ihrem Chefarzt vorgesetzt, was dem ja auch eine gewisse Legitimität mitgibt. Und wir hatten es

inzwischen Jahre her, da waren die KI-Systeme noch viel schlechter. 2018, 2019 mit Professor Dewey von der Charité-Untersuchung im radiologischen Bereich gemacht. Und da waren viele der Menschen, die selber die Befundung machen, die gesagt haben, naja, wer bin ich denn, um zu sagen, das System liegt falsch, weil es auch viel aufwendiger ist, meinem Chef, meiner Chefin zu erklären, warum das denn jetzt anders ist. Und im Zweifel bin ich dann, die, die haftbar ist. Und es ist viel einfacher zu sagen, das System hat das vorgeschlagen und ich schließe mich dem an. Einfach dieser ganze Prozess, dass ich das verargumentieren muss, dass ich im Zweifel dann dafür haftbar wäre, das ist ja auch immer ein großes, große Sorge bei Mediziner, Medizinerinnen. Und ich glaube, das ist etwas, wo, es kommt ja nicht nur aufs Finden drauf an, sonst kommt ja auch darauf an, zu sagen, ist das jetzt was, wo wir intervenieren müssen oder ist das was, was das System findet, aber was eigentlich irrelevant wäre. Und diese Abschätzung, dann zu sagen, das System hat was gefunden, aber ich sage, das ist irrelevant, die ist unglaublich schwierig, glaube ich, in dieser komplexen Verantwortungssituation. Ja, ja, total. Also, ja, widersprechen von so einem Systemvorschlag ist, glaube ich, so, dass, ich glaube, es ist wahrscheinlich einfacher zu sagen, das System hat nichts gefunden, ah, da ist doch was. Ich glaube, diese Entscheidung ist einfacher als zu sagen, oh, das System hat was gefunden und das ist dann irgendwie, das ist Quatsch. Ja, es ist ein total komplex und ich glaube, das ist auch eine total systemische Komplexität dahinter. Also wer macht die Diagnose, zu welchem Zeitpunkt? Das hat ja auch noch immer einen Rieseneinfluss. Ja, bin ich irgendwie, keine Ahnung, am Ende meiner 24-Stunden-Schicht und eh schon fertig? Oder bin ich jetzt gerade irgendwie wach und kann so was noch machen? Also, ja, es stimmt so, das ist ein ganz, ganz komplexes Flecht aus Verantwortungszuschreibung, Organisationsstruktur, Systemgenauigkeit, individuelle Expertise und Fähigkeiten, die dann am Ende in der Praxis dazu führen, dass, ja, wie da eigentlich mit so einem System umgegangen wird. (...) Gibt es noch eine letzte dringende Frage? Das können wir ja dann auch. Nein, da gibt es eine dringende Frage. Ich habe wo? Ich habe es nicht gesehen, wirklich? Ich sehe es nicht da. Ja, ich habe da noch die Frage, wie das mit der Erklärbarkeit ist und ob es

denn eine Erklärbarkeit gibt, wenn die Systeme Signale nutzen, die wir als Experten vielleicht gar nicht haben. Aber in der Radiologie werden ja zum Teil Filter beim Anlernen der Modelle benutzt. Die habe ich in meinem System gar nicht. Das heißt, das System oder die KI kann mir gar nicht erklären, wie sie zu der Entscheidung gekommen ist und wie gibt es da Daten zu oder Ideen zu, wie man denen begegnet? Also wir sprechen von, wenn wir von Erklärungen von solchen Modellen sprechen, von oft sogenannten Post-Hoc-Erklärungen. Also das System wird Ihnen nicht sagen, also Sie könnten auch gar nichts damit anfangen, wenn es jetzt Ihnen, ich zeige Ihnen, keine Ahnung, wenn es jetzt bei jedem dieser Knoten die einzelnen Parameter Ihnen zeigt, es ist ja keine Erklärung, es bringt ja nichts. Das heißt, was man macht, ist einfach von Post-Hoc nochmal einen zweiten Algorithmus drüber zu legen und zu sagen, zum Beispiel, das da ist ja de facto eine für einen Menschen wahrnehmbare Erklärung. Ja, irgendwie das ist da und da und da gefunden. Aber das ist einfach nochmal mehr oder minder ein zweiter Algorithmus, der nochmal drüberläuft, um zu zeigen, okay, wie vielleicht ist es auf ein Output gekommen? Also es sind Post-Hoc-Erklärungen. Also die erklären jetzt nicht, wie dieses neuronale Netz durchgelaufen ist, sondern eigentlich nur, wie das Ergebnis zustande gekommen ist, also wie man das Ergebnis interpretieren könnte, sagen wir es mal so. Gut, ganz herzlichen Dank, Elke Reine.