



Rechtsausschuss

18. Sitzung (öffentlich)

13. Juni 2023

Düsseldorf – Haus des Landtags

13:30 Uhr bis 14:56 Uhr

Vorsitz: Dr. Werner Pfeil (FDP)

Protokoll: Sitzungsdokumentarischer Dienst

Verhandlungspunkt:

Einsatz von ChatGPT im Justizbereich

3

Vorlage 18/1022

– Anhörung von Sachverständigen (*s. Anlage*)

* * *

Vorsitzender Dr. Werner Pfeil: Ich begrüße alle ganz herzlich zur 18. Sitzung des Rechtsausschusses. Ich begrüße die anwesenden und zugeschalteten Ausschussmitglieder, die hier im Saal anwesenden und zugeschalteten Sachverständigen, die Vertreterinnen und Vertreter der Landesregierung, alle Zuhörerinnen und Zuhörer, die auch online zugeschaltet sind, weil die Sitzung und die Anhörung online übertragen werden, alle Medienvertreter, die hier im Saal und online zugeschaltet sind und den Sitzungsdokumentarischen Dienst zu unserer heutigen Sitzung.

Der Ausschuss wurde mit Einladung 18/362 vom 31. Mai 2023 zu dieser Sitzung eingeladen, ein Neudruck erfolgte am 2. Juni 2023. Zu der Einladung liegen bisher keine weiteren Anmerkungen vor, sodass ich davon ausgehen kann, dass wir heute in die Tagesordnung direkt einsteigen können. Ich möchte darauf hinweisen, dass die komplette Anhörung heute per Livestream im Internet übertragen wird und anschließend als Video auch abrufbar ist. Änderungswünsche für die heutige Tagesordnung sind bisher nicht eingegangen. Gibt es Änderungswünsche von Ihnen? – Das sehe ich nicht. Dann darf ich die Tagesordnung so feststellen.

Ich rufe daher den einzigen Tagesordnungspunkt auf:

Einsatz von ChatGPT im Justizbereich

Vorlage 18/1022

– Anhörung von Sachverständigen (*s. Anlage*)

Die Fraktion der FDP hat am 9. März 2023 einen schriftlichen Bericht der Landesregierung zu diesem Thema angefordert. Dieser wurde als Vorlage 18/1022 in der Sitzung vom 22. März 2023 beraten. Diesen Bericht haben wir zum Anlass genommen, den Ihnen vorliegenden Fragenkatalog zu erarbeiten.

Mit Schreiben des Landtagspräsidenten vom 3. Mai 2023 wurden die Sachverständigen zur heutigen Anhörung geladen. Die Sachverständigen begrüße ich noch einmal ganz herzlich und freue mich, dass diejenigen, die heute hier persönlich anwesend sind, hier sind und bei den zugeschalteten freue ich mich, dass sie an der Anhörung am Online-Stream teilnehmen. Ich bedanke mich bei allen Sachverständigen jetzt schon für die sehr ausführlichen Sachverständigengutachten, die wir erhalten haben und mit denen wir uns auseinandersetzen konnten. Deswegen ein Hinweis von mir an dieser Stelle: Es wird gleich kein Statement zu Ihrem Sachverständigengutachten gefordert oder gewünscht, sondern es wird ganz konkrete Fragen der jeweiligen Abgeordneten der einzelnen Fraktionen geben, die noch offengeblieben sind oder bei denen noch weiterer Klärungsbedarf besteht.

Die Stellungnahmen und die jeweiligen Sachverständigengutachten liegen auch noch vorne im Eingangsbereich aus. Sie können eingesehen und auch mitgenommen werden.

Wir beginnen mit den Fragerunden, und der Antragsteller kann die erste Frage stellen. Danach folgen die CDU, die SPD, die Grünen und dann die AfD. Es ist so geplant, dass jede Fraktion maximal drei Fragen stellen kann. Die Fragen richten sich entweder

an alle oder einzelne Sachverständige. Wenn alle Fraktionen ihre Fragen gestellt haben, beginnen wir mit der Antwortrunde. Das heißt, die Sachverständigen schreiben sich die Fragen auf, die an sie jeweils gestellt wurden, und beantworten sie der Reihe nach. Soweit so gut. Alles verstanden? Gut.

Als antragstellende Fraktion darf ich die erste Frage an alle Sachverständigen stellen; denn alle Sachverständigen haben sich dahingehend geäußert, dass Urteile von Menschen abgefasst werden müssen, gleichzeitig haben wir aber auch Beispiele, wie insbesondere im Gutachten von Frau Biallaß gesehen, dass in anderen Teilen der Welt ChatGPT von Richtern zur Urteilsfindung oder Urteilsabfassung aber schon genutzt wird.

Die Frage, die sich dann für uns stellt: Was müssen wir oder was können wir als Landtag, als Exekutive bzw. als Justizausschuss zum Schutz des einzelnen vor einer falschen KI-Anwendung durch die Richterschaft durchführen, oder was müssen wir hier machen? Konkret regt Frau Biallaß an, die Exekutive solle sich zurückhalten und man sollte es der Justiz überlassen.

In dem Zusammenhang bitte ich Sie dann die Frage zu beantworten: Ist das bei einem so zentralen Thema, was die dritte Gewalt direkt und unmittelbar angeht, richtig? Benötigen wir hier nicht einen gesamtgesellschaftlichen Dialog? Müssen wir darüber nachdenken, ob es ausreichend ist, dass die Justiz selber aufgrund des Berufsethos möglicherweise eigene Richtlinien erlässt, oder benötigen wir Vorgaben des Justizministeriums? Benötigen wir unter Umständen gesetzliche Konkretisierungen auf Bundesebene oder ein Verfassungsgrundrecht in unserer Landesverfassung? Also: Was müssen wir als Exekutive machen, um den einzelnen vor einer falschen KI-Anwendung durch die Richterschaft zu schützen, oder reichen die bisherigen Regelungen? – Das wäre die Frage vonseiten der FDP-Fraktion, die sich an alle Sachverständigen richtet.

Jetzt gebe ich das Wort weiter an Frau Erwin für die CDU.

Angela Erwin (CDU): Vielen Dank, Herr Vorsitzender! Liebe Kolleginnen und Kollegen! Verehrte Sachverständige! Seitens der CDU-Fraktion möchte ich zunächst meinen Dank an Sie, liebe Sachverständige für die umfangreichen schriftlichen Stellungnahmen, die Sie uns zu diesem Thema zur Verfügung gestellt haben, richten. Auch herzlichen Dank sowohl an die Sachverständigen, die heute in Präsenz bei uns sind, es ist heute ein bisschen wärmer in diesem Raum, aber auch herzlichen Dank an diejenigen, die sich digital zugeschaltet haben.

Ich kann meine Frage an die meines Vorredners anschließen und sie allgemeiner an alle Sachverständige formulieren. Ich möchte wissen, warum Sie der Auffassung sind, dass ChatGPT zum Verfassen von Urteilen ungeeignet ist.

Meine weitere Frage richtet sich an Herrn Hartmann und Frau Biallaß. Sie haben in Ihren Stellungnahmen dargelegt, dass mit ChatGPT eine Reihe von Gefahren einhergehen. Sie haben dann ein paar Beispiele genannt. Könnten Sie uns das bitte näher ausführen, gerade auch vor dem Hintergrund und dem Kontext eines Richters, welche Gefahren Sie da sehen?

Sonja Bongers (SPD): Sehr geehrter Herr Vorsitzender! Verehrte Damen und Herren! Auch seitens der SPD-Fraktion ein ganz großes Dankeschön an alle Sachverständigen. Herzlichen Dank, dass Sie sich heute auch noch einmal die Zeit nehmen, um unsere offenen Fragen zu beantworten. Ich möchte direkt mit einer Frage starten, die ich sowohl Herrn Professor Dr. Heetkamp als auch Frau Biallaß stellen möchte. Sie beschreiben beide in Ihren Stellungnahmen die Bedeutung von LLM, also von diesen Large Language Models, und Sie beschreiben dort konkret, dass es bereits gewinnbringenden Einsatz von KI-Sprachtools in der Justiz gibt. Können Sie etwas klarer verdeutlichen, inwiefern diese konkrete Einsatzfähigkeit dieser KI-Systeme überprüft wird und wie man es sich vorstellen kann oder muss, dass zulässige Sprachtools für die Justiz weiterentwickelt werden?

Dann habe ich noch eine allgemeine Frage, die ich gerne allen stellen würde. Wir haben lesen können, dass in anderen Ländern teilweise die Anwaltschaft ChatGPT oder andere Produkte anwendet, ohne dass es dort Regelungen gibt. Dort ist es zu diversen Zwischenfällen gekommen, sodass die Justiz eigentlich erst einmal alles überprüfen muss, was die Anwaltschaft, was die Anwälte geschrieben haben. In konkreten Fällen ist es dazu gekommen, dass die Justiz völlig blockiert wurde, dass die entsprechenden Schriftsätze wirklich erst einmal tagelang auf Wahrheitsgehalt und die Verweisung auf entsprechende Urteile hin geprüft werden müssen. Können Sie sich vorstellen, dass so etwas in Deutschland passieren kann?

Vorsitzender Dr. Werner Pfeil: Für die Grünen hätte jetzt Frau Hanses die Möglichkeit, Fragen zu stellen. Sie ist noch nicht da, die Fragen werden dann später gestellt. Für die AfD dann Herr Professor Zerbin.

Prof. Dr. Daniel Zerbin (AfD): Sehr geehrter Herr Vorsitzender! Meine Damen und Herren! Ich möchte mich auch erst einmal recht herzlich bei den Gutachtern für die ausführlichen Stellungnahmen bedanken und dass Sie Zeit gefunden haben, entweder im Chat oder hier vor Ort in den etwas warmen Landtag zu kommen. Herzlichen Dank dafür.

Ich habe zwei Fragen. Eine Frage geht an Frau Biallaß, und dann an Herrn Professor Grabmair. Zur ersten Frage an Frau Biallaß. Sie führen in Ihrer schriftlichen Stellungnahme in der Antwort zur Frage 1 aus, dass der Einsatz von KI zur Assistenz bei der Entscheidungsfindung problematisch ist, wenn ihre Entscheidungsstruktur für den Richter nicht nachvollziehbar ist. Vor Kurzem forderten mehr als 1.000 Experten aus Tech und Forschung, unter ihnen auch Elon Musk, eine Pause bei der KI-Entwicklung. Es ging um ein sechsmonatiges Moratorium. Die Zeit solle genutzt werden, um das Regelwerk für die Technologie zu schaffen, was es so ja noch nicht gibt. KI-Systeme mit einer Intelligenz, die Menschen Konkurrenz machen, können größere Risiken für Gesellschaften und Menschheit bringen – wir haben gerade ein Beispiel gehört –, heißt es in dem offenen Brief der Befürworter einer Entwicklungspause und sogenannte generative KI wie ChatGPT-4 oder DALL-E sei mittlerweile so weit fortgeschritten, dass selbst die Entwickler ihre Programme nicht mehr verstehen oder wirksam kontrollieren

könnten, heißt es da weiter. Das sehen wir jetzt auch wieder in den Medien, es wird immer wieder berichtet.

Die Frage ist: Wie beurteilen Sie die Forderung nach einer Entwicklungspause? Herr Professor Grabmair, die Frage nach einer Entwicklungspause möchte ich auch an Sie adressieren. Sie führen in Ihrer schriftlichen Stellungnahme zu Frage 17 aus, dass die Erklärbarkeit von tiefen neuronalen Netzen ein in weiten Teilen ungelöstes Problem und Gegenstand aktiver Forschung ist. Sie sagen wörtlich:

„Nach meinem bestem Wissen ist unbekannt, auf welchen genauen Datensätzen die OpenAI Modelle ChatGPT und ChatGPT-4 trainiert wurden.“

Wie sehen Sie so eine Entwicklungspause für sich?

Vorsitzender Dr. Werner Pfeil: Herr Professor Zerbin, vielen Dank. – Frau Hanses ist jetzt da und darf auch drei Fragen an unterschiedliche Sachverständige stellen.

Dagmar Hanses (GRÜNE): Herr Vorsitzender, vielen Dank. – Liebe Kolleginnen und Kollegen! Sehr geehrte Sachverständige! Vielen Dank für Ihre Stellungnahmen. Das ist für mich ein Bereich, bei dem ich doch schön häufiger nachschlagen muss. Ich muss mir auch einen Überblick über die Fragen, die die Kolleginnen und Kollegen schon gestellt haben, verschaffen. Wir haben in verschiedenen Stellungnahmen gelesen – bei den Anregungen und Empfehlungen auf EU-Ebene oder auch auf nationaler Ebene –, dass man weiter am Programm forschen und diese Forschung empirisch untersucht werden sollte. Deshalb möchte ich meine Frage an Frau Biellaß, Herrn Professor Grabmair und Herr Nink wenden. Bitte korrigieren Sie mich, Herr Vorsitzender, wenn diese Frage bereits gestellt wurde.

Die Frage lautet: Bei ChatGPT kann man nicht nachvollziehen, was bei einer Antwort aufgrund von Quellen oder aufgrund von Wahrscheinlichkeit der KI selber generiert wurde. Wenn man das nicht nachvollziehen kann, muss man dann eigene Systeme schaffen, um das nachvollziehen zu können?

Vorsitzender Dr. Werner Pfeil: Die Frage wurde noch nicht gestellt. Das nur zur Klarstellung. Entsprechend des Tableaus möchte ich jetzt die Sachverständigen einzeln aufrufen und bitten, die an Sie gerichteten Fragen zu beantworten. Herr Professor Johannisbauer beginnt.

Prof. Dr. Christoph Johannisbauer (Hochschule für Polizei und öffentliche Verwaltung Nordrhein-Westfalen [per Video zugeschaltet]): Herzlichen Dank für die Benennung und für die Möglichkeit der Stellungnahme sowie für die Einladung zur heutigen Anhörung. Ein paar Fragen sind an alle gegangen und so auch an mich. Die erste Frage, die ich mir notiert habe, war, was die Exekutive machen kann, ob es Vorgaben für die Justiz geben soll und ob da Bedarf gesehen wird. – Ich würde ehrlicherweise die richterliche Unabhängigkeit als Erstes nach ganz vorne setzen. Es gibt ein hohes Berufsethos, es gibt Disziplinarvorschriften. Ich würde es nicht so wie einen Negativkatalog sehen, wie: Was darf die Richterschaft nicht? Darf sie überhaupt ChatGPT

einsetzen? Ich würde keinen Negativkatalog aufsetzen nach dem Motto, was sie alles nicht machen darf. Davon halte ich ehrlich gesagt ganz allgemein nichts. Das war der erste Punkt.

Die zweite Frage war etwas allgemeiner. Warum ist ChatGPT aus Sicht der Sachverständigen, also aus meiner Sicht, zur Entscheidungsfindung ungeeignet? Sie haben es vielleicht meiner Stellungnahme entnommen, dass sich das alles etwas systematisieren wird. Ich denke, alle Sachverständigen waren einer Meinung, dass ChatGPT oder KI an sich nicht Richter sein kann, das können nur natürliche Personen, also Menschen sein. Das sagt uns die Verfassung. Als Werkzeug könnten LLM-Systeme wie ChatGPT eventuell genutzt werden. Da muss man sich genau ansehen, auf welcher Stufe wir uns befinden. Wenn man es systematisieren möchte, muss man immer drei Fragen stellen: Ist es technisch möglich, ist es sinnvoll und ist es rechtlich zulässig?

Schauen wir uns zum Beispiel den Punkt, Entscheidungsvorschläge zu unterbreiten, an. – Es geht nicht um einfache Assistenzaufgaben oder um die Erstellung einfache Schreiben. – Die Entscheidung selbst zu finden, geht nicht, da haben wir einen Verstoß gegen das Verfassungsrecht. Entscheidungsvorschläge zu machen oder Teile von Entscheidungen vorzuschlagen, bringt auch Probleme: Erstens ist es technisch aktuell nicht möglich. Ich habe dazu verschiedene Fälle und Fragestellungen aus der Lehre testweise versucht. Dabei kamen nicht wirklich gute Dinge heraus. Die Beantwortung von Fragen juristischer Laien funktioniert noch einigermaßen, wenn es aber wirklich um Fachfragen geht, funktioniert es nicht. Bei praktischen [akustisch unverständlich], das hat auch die Stellungnahme der Richterschaft gezeigt, funktioniert es mit ChatGPT überhaupt nicht. Technisch ist es aktuell nicht wirklich möglich, das hat auch etwas mit der Art, wie es konzipiert ist, mit Wahrscheinlichkeitsberechnung zu tun. Dann kommt ein ganz großer Punkt, nämlich die Frage, ob technisch als auch sinnvoll ... Der menschliche Richter ist ja kein Subsumtionsautomat: Es kommen oben die Tatsachen herein, und unten das Ergebnis heraus. Es sind ja noch viel mehr Dinge zu berücksichtigen, wie ethische, soziale und moralische Aspekte, die einfach nicht von der KI abgebildet werden können.

Allein aus diesen grundsätzlichen Überlegungen heraus, ist es derzeit nicht denkbar, dass ChatGPT eingesetzt wird. Bei ChatGPT kommen noch weitere Dinge hinzu wie datenschutzrechtliche Risiken, Manipulationsgefahr durch den Betreiber. Es müssen ja auch Trainingsdaten eingegeben werden. Es wird ja auch trainiert: Was ist erwünscht, was nicht? Da besteht eine große Manipulationsgefahr. Dann haben wir fehlende Transparenz. Hier haben wir das Stichwort Blackbox und die ganzen Probleme, die in der Beeinflussung der Richter liegen, durch ein solches System, Stichwort Automatic Vice. Von daher ist ChatGPT für die Entscheidungsfindung nicht zu gebrauchen. – Das war die zweite Frage.

Die dritte Frage war: Kann das die Justiz blockieren, wenn Anwälte alle ChatGPT einsetzen? So soll es schon passiert sein. – Ja. Das kann ich mir sehr gut vorstellen. Das erinnert mich ein bisschen an Fälle, in denen Reichsbürger zum Beispiel zu Gericht gehen. Da ist es gang und gäbe, sich aus dem Internet, 60, 70, 80 Seiten Kauderwelsch herunterzuladen und das bei Gericht einzureichen. Genau da stellt sich die

Frage, ob der Richter das alles lesen muss. – Ja, wohl schon. Das ist schon potenziell geeignet, die Justiz zu blockieren. So muss man es wohl sagen.

Dr. David Nink: Herzlichen Dank für die Gelegenheit zur Stellungnahme und für die Möglichkeit, in die Diskussion einsteigen zu können. Bei der ersten Frage möchte ich mich gerne meinem Vorredner anschließen. Aus meiner Sicht ist das Richterethos durchaus ernst zu nehmen. Die Eigenwahrnehmung der Richterinnen und Richter als hohes Gut und von Verfassungsrang, die richterliche Unabhängigkeit insbesondere in Form der sachlichen Unabhängigkeit, woraus ein bisschen folgt, was eigentlich zu tun ist. In den Kernbereich der Rechtsfindung seitens der Exekutive einzugreifen, ist ohnehin schwierig. Ich könnte mir vorstellen, dass es denkbar ist, dass so etwas wie ein Recht auf den menschlichen Richter irgendwann einmal kodifiziert wird, und zwar dann, wenn die technischen Möglichkeiten solcher Systeme noch weiter steigen oder vom Bundesverfassungsgericht hergeleitet werden. Das haben wir beispielsweise in der Vergangenheit zum Grundrecht zur Integrität informationstechnischer Systeme hin und wieder schon gesehen. Das steht nicht ausdrücklich in der Verfassung, wird aber vom Bundesverfassungsgericht entsprechend hergeleitet. Das wäre theoretisch auch mit steigenden technischen Möglichkeiten denkbar.

Die zweite Frage von der CDU-Fraktion war: Warum ist ChatGPT eigentlich ungeeignet? – Ganz profan mal vorangestellt: Bei den Nutzungsbedingungen von ChatGPT schreibt der Betreiber OpenAI ausdrücklich, dass dieses System nicht geeignet ist, es im Bereich der Gesundheitsvorsorge und auch im Bereich der Rechtsanwendung und Rechtsberatung heranzuziehen. – Das klingt zwar profan, zeigt aber, dass selbst der Betreiber davon ausgeht, dass das Tool dafür nicht geeignet ist. Weiterhin sprechen natürlich datenschutzrechtliche Bedenken dagegen. Hier haben wir ein System, das in den USA gebaut, gehostet und betrieben wird. Selbst wenn man keine Justizinterna in diese Prompts eingeben würde, hätte man doch immer die Gefahr des Datenschutzrechts, sobald Rückschlüsse auf natürlich Personen möglich sind und wir dann im Bereich personenbezogene Daten wären.

Ein weiterer Punkt ist, dass die Ergebnisse im Moment auch qualitativ noch nicht gut genug sind, um den Standard zu erreichen, den menschliche Richterinnen und Richter erreichen, insbesondere um menschliche und juristische Lebenssachverhalte adäquat abbilden zu können.

Mein Vorredner ist ja schon auf die Unterscheidung zwischen Entscheidungsfindung und reines Ausformulieren eingegangen. Jedenfalls zur Entscheidungsfindung und Entscheidungsvorbereitung ist es technisch derzeit noch nicht geeignet. Wenn man großflächig menschliche Lebenssachverhalte in einen Urteilsstil packen will, ist das derzeit qualitativ auch noch nicht geeignet, liegt aber vielleicht auch ein bisschen daran, dass das System natürlich nicht ausdrücklich auf die deutsche Rechtsanwendung hin trainiert und optimiert wurde. Darauf kommen wir vielleicht im Verlauf noch zu sprechen. Herr Professor Grabmair wird eventuell noch etwas dazu sagen. Wenn man aber so ein Modell auf das deutsche Recht hin trainiert und optimiert, könnte man erwarten, dass dann auch geeignetere Ergebnisse herauskommen.

Ein weiterer Punkt ist, dass auch die Transparenz aus meiner Sicht nicht gegeben ist. Ich habe ein wenig ausprobiert und soweit es geht, mich aus öffentlichen Quellen zur Funktionsweise und Datenbasis informiert. Es gibt Beschreibungen zur Datenbasis, aber keine genaue Auflistung, welche Texte für das Training und die Validierung genutzt wurden. Letztlich ist auch die Funktionsweise für eine konkrete Antwort für den Anwender nicht nachvollziehbar. Man kann ein bisschen versuchen, sich die allgemeine Funktionsweise mit technischem Verständnis, mit Recherche herzuleiten, aber die genaue Funktionsweise, wie man von einer Frage zu einer Antwort kommt, ist für den durchschnittlichen Anwender meines Erachtens nicht gegeben.

Die nächste Frage von der SPD-Fraktion würde ich ganz knapp bejahen. Die Gefahr, dass die Justiz blockiert wurde, sehe ich durchaus auch. Dann ist ein valider Gedanke zu überlegen, ob die Justiz auch „aufrüsten“ muss, um solche großen Datenmengen und riesigen Schriftsätze, wie sie jetzt in Massenverfahren vorkommen, bewältigen zu können. Dass man quasi Schritt hält, ist kein Selbstzweck, aber meines Erachtens ein durchaus valider Gedanke.

Die letzte Frage kam von der grünen Fraktion und bezog sich, wenn ich das richtig notiert habe, auf die Nachvollziehbarkeit. Sollte man ein eigenes System in Deutschland schaffen, um die Nachvollziehbarkeit zu erhöhen? – Ja, das sollte man meines Erachtens; denn die Nachvollziehbarkeit jeglichen staatlichen Handelns hat Verfassungsrang. Trainingsdaten sind nicht einsehbar, die Trainingsmethodik kann man so halb nachvollziehen, aber nicht im Detail, wie man zu einer konkreten Antwort kommt. Auch die Optimierung hin auf deutsches Recht spielt in den Bereich Nachvollziehbarkeit hinein.

Ich weiß nicht, ob wir später noch dazu kommen, aber ChatGPT wurde unter anderem auf das amerikanische Bar Exam trainiert. Das hat einen anderen Prüfungsgegenstand und eine andere Prüfungsmethodik als die deutsche Juristenausbildung; Prüfungsgegenstand in dem Sinne, dass schon das US-amerikanische Recht auf dem Commonlaw beruht und dass es eben eher ein korrelativer Rechtsfindungsprozess ist. Man orientiert sich an ähnlichen Fällen. Das kommt solchen Sprachmodellen eher entgegen als das kodifizierte deutsche Recht. Zum anderen ist die Juristenausbildung bei uns eine ganz andere. Im US Bar Exam gibt es auch Multiple-Choice-Fragen und dergleichen. Dann kommt es eher darauf an, sich an Fälle zu erinnern oder Fälle zu finden, die so ähnlich sind. Das heißt, das ist nicht ganz vergleichbar. Daher meine Antwort auf die Frage: Ja, man sollte eigene Systeme schaffen, um die Nachvollziehbarkeit zu erhöhen, und dann natürlich auch umfassend verschiedene Disziplinen daran beteiligen.

Isabelle Biallaß (Deutscher EDV-Gerichtstag): Ich möchte mich auch zunächst für die Gelegenheit zur Stellungnahme bedanken. Jetzt haben meine Vorredner zu den an alle gerichteten Fragen schon so viel gesagt. Die erste Frage, ob die Exekutive tätig werden muss, um sicherzustellen, dass Urteile in Zukunft von Menschen gefertigt werden, habe ich ja schon in meiner schriftlichen Stellungnahme beantwortet. Das war ja auch Anlass dafür, warum diese Frage jetzt gestellt wurde.

Ich bleibe dabei, dass das notwendig ist, um die richterliche Unabhängigkeit zu wahren. Nichtsdestotrotz kann meiner Meinung nach Unterstützung erfolgen, nämlich

durch den von mir in der Stellungnahme angesprochenen Punkt, „wir brauchen mehr sogenannte Digital Literacy“, also Digitalkompetenz bei den Technologien zukünftig einsetzenden Richterinnen und Richtern sowie Staatsanwältinnen und Staatsanwälten. Es muss sichergestellt werden, dass es entsprechende Fortbildungsangebote gibt, dass motiviert wird, dass man sie einsetzt. Denn nur, wenn ihnen bewusst wird, wo die Grenzen derartiger Systeme liegen, kann sichergestellt werden, dass sie ihr Amt in Zukunft noch angemessen ausüben. Da sehe ich tatsächlich auch einen gewissen Tätigkeitsbereich der Exekutive, dass sie sicherstellt, dass die entsprechenden Formate geschaffen werden, möglicherweise auch kurzfristig.

Wir haben als Richter die Angewohnheit, sehr viel im Homeoffice zu arbeiten. Bis alle Geräte gestellt bekommen haben, war es durchaus üblich, die eigenen Geräte dort einzusetzen. Ich sehe die Gefahr, dass gerade Kollegen, die sich nicht so stark mit den technischen Hintergründen befassen und ein gefährliches Halbwissen haben, plötzlich in die Versuchung kommen, Large Language Models für ihre Arbeit einzusetzen. Dem müssen wir entgegenwirken.

Damit komme ich zu der zweiten Frage, nämlich zu der Frage: Warum ist der Einsatz von ChatGPT in der Justiz nicht möglich? Ist es ungeeignet, um auf dieser Grundlage Entscheidungsvorschläge erstellen zu lassen? – Da hat mir die sehr ausführliche Antwort von Herrn Dr. Nink alle Punkte, die ich auch schon in meiner schriftlichen Stellungnahme aufgelistet habe, genommen. Es beginnt, wie er schon ausgesprochen hat, mit den Allgemeinen Geschäftsbedingungen von OpenAI, weiter über datenschutzrechtliche Bedenken und dann auch um das Problem, wie denn das Programm überhaupt technisch ausgestaltet ist. Das Problem an der guten Ausdrucksweise ist, dass bei dem nicht technisch vorgebildetem Nutzer der Eindruck entsteht, wenn ich das System bitte, ein Urteil zu erstellen und oben ein Tenor steht und unten eine Begründung, dass der erstellte Tenor tatsächlich auf der untenstehenden Begründung fußt. Das ist aber überhaupt nicht der Fall. Man hört immer so häufig „explainable AI“, und dann kommen Personen in Versuchung, zu denken, das Programm erklärt mir doch, warum es den Tenor geschrieben hat. Das ist aber technisch überhaupt nicht das, wie es funktioniert. Hier bin ich wieder an dem Punkt, an dem wir eine entsprechende Awareness schaffen müssen. Die weiteren Aspekte, weil ich mich ansonsten wiederholen würde, würde ich an dieser Stelle übergehen.

Ich hatte in meiner schriftlichen Stellungnahme darauf hingewiesen, dass es bereits Einsätze von Large Language Models in der Justiz, insbesondere durch Rechtsanwälte gibt, und dass es KI-Einsatz grundsätzlich auch schon in der deutschen Justiz gibt.

Da muss man differenzieren. Mir ist noch kein Einsatz in Deutschland von Large Language Models bekannt. Wir haben diverse Pilotprojekte, die andere KI-Formen einsetzen. Beispielsweise wurde beim Amtsgericht Frankfurt zum Einsatz bei der Bearbeitung der Klagen von Fluggastrechten das Programm „Frauke“ von IBM entwickelt erprobt, ebenfalls KI-basiert ist „OLGA“ beim OLG Stuttgart im Einsatz. Beide helfen bei der Bearbeitung von Massenverfahren. Dann gab es Pilotprojekte einer Software namens „Codefy“. Die wurde beim Landgericht in Ingolstadt pilotiert und wird, soweit ich weiß, aktuell pilotiert beim Landgericht in Hechlingen.

Wir haben unseren E-Akten-Hersteller E2A von Sinc. Unsere jetzt im Einsatz befindlichen Softwareanwendungen enthalten schon Bereiche zur Durchdringung, je nachdem wie man den KI-Begriff fasst, möglicherweise unter KI subsumieren könnte, aber kein aufwendiges Machine-Learning. Aber Sinc hat auch Lösungen, die tatsächlich einen höheren KI-Einsatz haben, im Angebot und werden unter dem Begriff „Jute“ vermarktet. Auch das könnte man theoretisch, wenn sich der Bund dafür entscheiden würde, bei uns integrieren. Ich weiß, in Niedersachsen gibt es im Augenblick Planungen, ein entsprechendes Forschungsprojekt dazu anzustrengen. Bei dem Einsatz bei der Staatsanwaltschaft – da will ich Herrn Hartmann nicht vorgreifen – gibt es auch diverse KI-Einsatzszenarien. Man sieht also: Es gab schon, bevor dieser große Trend zu LLM kam, ganz viele Überlegungen.

Jetzt haben wir ein Werkzeug, das sprachlich viel mächtiger ist. Da wird nun der nächste Schritt sein zu überlegen, wie man diese viel größeren Fähigkeiten auch für die Justiz fruchtbar machen kann. Die Anwaltschaft schläft da nicht. Es gibt jetzt schon auf dem deutschen Markt einen Anbieter namens JUNE, bei dem man schon die Software von OpenAI nutzen kann. Es gibt weitere Anbieter, die damit werben, dass sie in Kürze auf den Markt gehen. Auf dem internationalen Markt gibt es Anwendungen, die noch deutlich mehr Komponenten anbieten, als das, was wir auf dem deutschen Markt haben. Man kann sich ein bisschen vorstellen, was auf Anwaltsseite in Zukunft passiert. Da sollten wir die Entwicklung nicht verschlafen, sondern überlegen, wie wir sie für uns in der Justiz nutzbar machen können.

Der nächste Punkt betrifft genau diesen Aspekt, den Einsatz von ChatGPT in der Anwaltschaft, dass es in anderen Ländern schon Zwischenfälle gab, dass Schriftsätze durch ChatGPT erstellt wurden und dann aufwendige Überprüfungen sämtlicher Fundstellen durch die Justiz notwendig waren. Man muss dazu sagen, bei uns in Deutschland wäre das medial gar nicht bekannt geworden, wenn das bei uns auch passiert wäre, durch das unterschiedliche Vorgehen. Zum einen ist es bei uns sowieso so, dass ich als Richter die Fakten aus dem Sachverhalt herausarbeite und im Anschluss die Rechtsfindung selber mache. Wenn ich dann ein Fehlzitat, was von einem Large Language Model erstellt wurde, habe, dann finde ich das nicht in der Datenbank. Dann wundere ich mich, wie es denn sein kann, aber „Quellen“ haben Studenten, Rechtsanwälte und andere Personen auch schon selbst erfunden. Dann würde ich denken, da hat jemand versucht, mir einen Beweis für seine Argumentation unterzujubeln, und tatsächlich gibt es den gar nicht. Das würde bei mir nicht als Erstes den Verdacht auslösen, er habe ein Sprachmodell genutzt.

Ich sehe aber die Gefahr, dass sich die Massenverfahren, die wir jetzt schon als große Herausforderung für die Justiz haben, noch stark multiplizieren, dass es einen gewissen Typ Anwalt gibt, der sagt: Ich habe jetzt eine tolle neue Möglichkeit, meine Schriftsätze schneller zu erstellen und noch mehr aufzublasen. Natürlich muss er aus anwaltlicher Sorgfalt gucken, dass er noch immer so mitwirkt, dass er alle relevanten Fakten mitteilt. Ob er aber diese Fakten wie bisher auf 20 Seiten oder in Zukunft auf 80 Seiten darlegt, weil er ein Large Language Model einsetzt, ist eine Entwicklung, die ich nur mit Sorge beobachten kann. Da muss man überlegen, wie man dem, wenn es diese Tendenz gibt, Herr wird. Das werden nicht alle sein, aber die Gefahr besteht natürlich.

Dann komme ich zur nächsten Frage, ob man die von Elon Musk und Co sechsmonatige Entwicklungspausieren befürworten sollte. – Ich kann jetzt nicht beurteilen, welche Motivation die Unterzeichner dieses Moratoriums hatten. Wenn man sich überlegt, dass zumindest Elon Musk parallel selber an der Entwicklung von Large Language Model arbeitet, mag man sich fragen, ob er nicht einfach nur sechs Monate Zeit haben wollte, um den Rückstand aufzuarbeiten und das gar nicht von dem Schutz der Welt vor irgendwelchen Bedrohungen getragen war.

Ich finde, wir sind in Europa da eigentlich schon ganz gut aufgestellt. Wir haben uns, bevor ChatGPT veröffentlicht wurde, mit dem AI-Act befasst. Daher hatten wir auch die Möglichkeit, jetzt relativ schnell auf die generativen Sprachmodelle zu reagieren und zu überlegen, wie wir das einbauen können. Es wird noch eine riesige politische Herausforderung werden, da eine Lösung zu finden, die einerseits die Innovation nicht hemmt, aber andererseits die Gefahren eingrenzt. Da hoffe ich, dass wir auf Europaebene zu einer guten Kompromisslösung kommen. Das ist meiner Meinung nach der richtige Weg, nämlich gesetzliche Regelungen zu erstellen, statt mit der Entwicklung aufzuhören. In sechs Monaten wird die Situation auch nicht wesentlich besser sein, sodass ich diese Aufforderung in keiner Weise mittragen oder unterschreiben würde.

Somit kommen wir zu dem Punkt entsprechende Erforschung der Einsatzmöglichkeiten von Large Language Models und die empirische Untersuchung ihres Einsatzes. Das kann ich so vollumfänglich befürworten. Ich bin der Meinung, dass das ein Punkt ist, mit dem wir uns in der Justiz befassen müssen. Wie sind die Chancen dieser Modelle? Für so etwas muss man dann auch mal Geld in die Hand nehmen. Ich habe ja gerade schon geschildert, dass man auf Anwaltsseite diese Technik nutzen wird. Wir müssen in der Justiz ein Konzept dafür entwickeln, wie wir damit umgehen wollen. Wir dürfen hierbei nicht in einen technischen Rückstand geraten.

Die letzte Frage ging im Prinzip in die gleiche Richtung und lautete: Müssen wir ein eigenes System schaffen, das sicherstellt, dass zu nachvollziehbaren Ergebnissen gekommen wird? – Meine ganz eindeutige Meinung dazu: Ja, das müssen wir. Wir haben hier eine sehr spezielle Domäne, wir haben als Justiz die Datenhoheit, über die Daten, die man für das Training eines solchen Modells braucht und wären dann in der Lage, Konzepte zu entwickeln, wie man das fruchtbar machen kann. Wir haben hier, und das haben wir gerade von Herrn Dr. Nink gehört, die zusätzliche Herausforderung, dass wir gerade kein Caselaw haben, sodass man auch Überlegungen anstrengen müsste, um unser Rechtssystem abzubilden, wie man gleichzeitig auch unser Recht abgebildet bekommt. Die ersten Versuche der Rechtsinformatikabbildung unseres Rechtssystems sind gerade an der Komplexität gescheitert. Für ausgewählte Fälle müsste man überlegen, ob man nicht eine Kopplung von Entscheidungsräumen hinkriegen würde. Wir stehen hierbei ganz am Anfang der Überlegungen, aber dieser Überlegungen sollten dringend hinreichend finanziert werden, damit wir sie auch in der Justiz selbstbestimmt durchführen können.

Prof. Dr. Matthias Grabmair (Technische Universität München, TUINF27 Informatik 19 – Professur für Legal Tech [per Video zugeschaltet]): Auch ich danke vielmals für die Gelegenheit zur Stellungnahme und zur Teilnahme an dieser Anhörung. Vielen

herzlichen Dank für die Einladung. Mir geht es ähnlich wie meiner Vorrednerin. Auch ich kann in großen Teilen nur auf das verweisen, was meine sehr guten Vorredner auch schon gesagt haben. Ich gehe trotzdem auf die einzelnen Fragen ein und versuche vor dem Hintergrund meiner Expertise, noch Inhalte hinzuzufügen.

Zu der ersten Frage, was die Exekutive gegenüber der Judikative hier tun kann/muss, möchte ich mich als zum Schwerpunkt technisch Arbeitender eigentlich zurückhalten, weil ich zu wenig tiefer Jurist bin und zu wenig die Kommunikations- und Arbeitskultur einschätzen kann, um hier eine Empfehlung auszusprechen. Das, was von den Kolleginnen und Kollegen erwähnt wurde, halte ich für sehr sinnvoll. Dies betrifft die richterliche Unabhängigkeit. Diese sollte vor dem Hintergrund der Motivation einer Technologieaufsicht auf keinen Fall in einer Art und Weise beschränkt werden, die man anderweitig als unzulässig empfinden würde.

Die erste Frage der CDU-Fraktion von Frau Erwin, warum wir glauben, dass die Modelle zur Erstellung von Urteilen ungeeignet sind, wurde schon von Herr Dr. Nink sehr gut beantwortet. Er hat beispielsweise auf die Datenschutzproblematiken, den Hinweis von OpenAI, dass eine Anwendung in der Justiz nicht angedacht wurde und davor gewarnt wird, hingewiesen.

Ich möchte hier hinzufügen, dass man, wie von Herrn Professor Johannisbauer angesprochen, zwischen der eigentlichen rechtlichen Entscheidungsfindung und der Entscheidungsvorbereitung und anderen administrativen und Rechercheaufgaben unterscheiden muss. Es geht auch aus den anderen Stellungnahmen hervor, dass es Tätigkeiten bei Gericht gibt wie das Vergleichen von Dokumenten, Teil der Recherche, tabellarische Aufstellungen usw. Da gibt es Tätigkeiten, die letzten Ende in der Pipeline zur rechtlichen Entscheidungsfindung sind, die sehr aufwendig wären, wenn sie händisch erledigt werden müssten, von Computersystemen aber schneller erledigt werden können und schnell validiert werden können, weil [akustisch unverständlich] der Zeitstrahl meinetwegen, der aus dem Dokument erstellt wird, jetzt automatisch dem entspricht, also Navigationshilfen in der E-Akte usw. Da gibt es auf jeden Fall Tätigkeiten, bei denen die Risikoprofile geringer sind als bei der Kernentscheidungsfindung. Hier lohnen sich Beschäftigungen mit den speziellen Funktionsweisen.

An der Stelle möchte ich hinzufügen, dass ich generell der Meinung bin, dass der Diskurs hier ein etwas höheres Niveau verdient an dem Punkt, als nur über den Einsatz von ChatGPT, über einzelne spezielle Funktionen und ihre Risiken zu sprechen. Hier darf ich eine kurze Analogie zum Product CoCounsel der amerikanischen Firma Casetext ziehen. Das ist eine Softwaresuite, die auf ChatGPT-4 basiert und für amerikanische Anwälte Funktionalitäten wie das Verfassen von Zusammenfassungen, kleine Rechercheaufgaben, tabellarischen Dokumentenvergleich usw. übernimmt. Das System ist eine klassische grafische Benutzeroberfläche, die unter der Haube mit einer ChatGPT-4-Version arbeitet. Das System erlaubt zwar dem Benutzer, einzelne Selbstprompts einzugeben, aber das wird mit einer Warnung versehen, wenn ich das richtig in Erinnerung habe: Liebe Benutzer, passen Sie auf, hier können Dinge kommen, die wirklich geprüft werden müssen.

Die Industrie hat schon erkannt, dass die Probleme, die sich durch die direkte Interaktion mit dem System ergeben, nämlich diese anthropomorphe-artige Illusion, dass ich

es mit einem Individuum zu tun hätte, das mir Wissen präsentiert, minimiert, aber zumindest ergänzt werden können, durch Funktionen, in denen nur mittels einer Oberfläche interagiert wird, die dem Benutzer das Prompten abnimmt. Das reduziert schon bestimmte Risikoprofile. Ich möchte jetzt nicht so verstanden werden, dass ich der Meinung bin, dass diese Systeme jetzt Richtern zwangsweise zur Verfügung gestellt werden sollten, aber dennoch befinden wir uns hier am Anfang einer technologischen Entwicklung auf der Basis dieser Foundation-Models, wie sie ja genannt werden, die ja noch weiter geht. Wir sollten dem auf jeden Fall pro aktiv und grundsätzlich technologieoffen begegnen.

In jedem Fall, und da bin ich völlig d'accord mit allem, was bisher gesagt wurde: Die Kernentscheidungsfindung muss bei dem menschlichen Richter bleiben. In diesem Fall ist das Risiko, dass durch eine unachtsame Nutzung dieses Systems Teile dieser Entscheidungsaspekte in dieses Advantage-System delegiert werden, als sehr kritisch zu betrachten und kann als solches nicht ermöglicht werden. Da teile ich die Vorsicht der anderen hier anwesenden Sachverständigen.

Die nächste Frage war von der SPD-Fraktion und bezog sich auf die Überschwemmung der Gerichte. Da erlaube ich mir kurz, zu abstrahieren. Das Problem haben wir im Prinzip seit dem Bestehen der Textbausteingeneratoren. Das gab es ja schon vor ChatGPT. Dementsprechend meine ich, dass vor diesem Hintergrund justizseitig weniger getan werden sollte als vorher. In dem Zusammenhang will ich davor warnen, dass die Sprachmodelle nur um ihrer Verwendung Willen zum Einsatz kommen.

Bei dem Fluggastrechteverfahren „Frauke“ in Frankfurt wird mit einer durch den Menschen getroffenen Entscheidung, die mit einem Formular spezifiziert wird und zu einem Urteilsbaukasten führt, der dann Textbausteine vorschlägt, ein System eingesetzt, das von Experten entwickelt wurde und völlig deterministisch agiert. Danach muss nichts mehr geprüft werden, weil kein Text mittels eines statistischen Algorithmus generiert wird. Es kostet Aufwand, das zu entwickeln, weil die Logik spezialisiert und die Textbausteine hergestellt und gewartet werden müssen, aber dafür bekommt man den Determinismus, der die Sicherheit bietet. Meiner Ansicht nach lohnt sich hier ein technologieoffener Umgang mit Large Language Models, allerdings auch nur als ein Baustein von vielen, mit denen man den Herausforderungen Herr wird.

Wenn jetzt anwaltsseitig diese Modelle dazu verwendet werden, vielmehr Klagen einzureichen und vielmehr – wie schon gesagt wurde – fiktive Zitate in diesen Schriftsätzen zu verwenden, muss ich als technisch Forschender kurz ergänzen, dass es die Fähigkeit von Sprachmodellen, fiktive juristische Zitate zu zitieren, schon länger gibt. Sie gibt es schon mindestens seit ChatGPT-2, seit Jahren sage ich mal, seit mindestens zwei Jahren ist es bekannt. Von daher ist das keine neue Situation, sondern nur ein Phänomen, was jetzt aufgetaucht ist. Auch hier bin ich der Meinung, dass das keine systemischen Gründe sind, sich der Benutzung dieser Systeme zu verschließen, sondern die Liste der Anforderungen, mit denen man nach technischen Lösungen sucht, zu erweitern.

Von Herrn Professor Zerbin kam die Frage nach der Entwicklungspause von großen KI-Systemen. Ich schließe mich Frau Biallaß an, dass die genauen Motive und Zwecke dieser Pause dahingestellt sein können. Ich möchte mir hier keine Empfehlung anmaßen,

dass die Forschung anzuhalten hat. Ich wünsche mir vielmehr, dass ein achtsamer Umgang, Einsatz und Untersuchung dieser Systeme stattfindet, speziell in den Fachkontexten, in denen es kritisch ist, wie hier in der Justiz. Dementsprechend sollte hier auch eine Demokratisierung stattfinden. Wir haben es ja mit Closed-Source-Systemen zu tun. Wir sehen jetzt gerade von anderen großen amerikanischen Firmen, dass deren Modelle und ihre Parameter zusehends in Open Source zur Verfügung gestellt werden, was ich sehr begrüße. Wir sind hier am Anfang einer Entwicklung, die mittelfristig dazu führen wird, dass wir dieses System sukzessive besser verstehen, gepaart mit einer größeren Technologieaffinität, die auf viele fachkundigere Anwender treffen wird. Diesen Weg sollten wir auf jeden Fall proaktiv gehen. Hier eine Pause zu fordern, löst unsere Probleme nicht, wenn wir stattdessen nicht proaktiv Dinge tun, die die Risiken minimieren und die Kompetenz stärken.

Ich muss ergänzen, dass der Einsatz dieser Modelle wie zum Beispiel Casetext, wie schon gesagt, nicht immer über einen Prompt stattfindet. Diese Modelle werden unter der Haube in den meisten, wenn nicht allen juristischen Dokumentanalyse-Systemen mittelfristig zum Einsatz in den E-Akten-Systemen usw. kommen. Die Modelle werden in die Suchpipelines integriert, sie werden in die Identitätsmarkierungskomponenten mit einfließen usw. Diese Modelle kommen, auch wenn der Nutzer nicht direkt mit ihnen in eine Prompt-Interaktion kommt.

Übergehend zur letzten Frage aus der grünen Fraktion. Wenn diese Modelle schon nicht nachvollziehbar sind, sollten wir dann nicht eigene Modelle bauen? – Die Antwort ist aus meiner Sicht eindeutig Ja. Die Erklärbarkeit, warum man bei der vorangehenden Frage noch angepasst ist, ist natürlich eine offene Forschungsfrage. Dennoch sehen wir, dass die Technologien, die vorhanden sind, neue Realitäten schaffen. Dementsprechend ist es an den Stakeholdern und den potenziellen Anwendern und den Anwendergruppen wie in der Justiz, wie Frau Biallaß schon sagt, ein Ökosystem zu schaffen, in dem eine nachhaltige achtsame Entwicklung in der Erklärbarkeit eine größere Rolle spielt und für alle gewährleistet wird.

Für mich führt das alles nicht zu einem Verbot, sondern zu einer proaktiven Beschäftigung. Dementsprechend glaube ich, dass wir in Deutschland vielleicht mit einer langsameren Entwicklung als in den USA und vielleicht mit kleineren Modellen als die der großen kommerziellen Anbieter, aber mit einer nachhaltigen dauerhaft kompetenzfördernden Vorgehensweise aus meiner Sicht sehr gut beraten sind. Ich sehe hier die Chance, dass die Justiz aus der reaktiven Position gegenüber eines kommerziellen Produkts zu einem aktiven Gestalter von was werden kann. Das Anforderungsprofil der Justiz gilt es noch zu etablieren und in eine nachhaltige Entwicklung einfließen zu lassen. In dem Sinne empfehle ich hier, eine proaktive Technologie [akustisch unverständlich] geleitet durch den Konsens. Von daher denke ich, dass die Entscheidung beim menschlichen Richter bleiben soll. Das empfehle ich als Vorgehensweise.

Markus Hartmann (Zentral- und Ansprechstelle Cybercrime Nordrhein-Westfalen): Herr Vorsitzender! Meine Damen, meine Herren! Vielen Dank für die Einladung. Die an mich gestellten Fragen beantwortet ich natürlich gerne. Die erste Frage war: Was kann der Landtag, was kann die Exekutive tun, um einen Rahmen zu bieten? – Ich

knüpfe gerne an das an, was meine Vorrednerin, meine Vorredner gesagt haben. Ich bin davon überzeugt, dass natürlich Dinge wie das Berufsethos der Richterschaft – hier erlaube ich mir, aus beruflicher Position ein bisschen zu erweitern – und der Staatsanwältinnen und Staatsanwälte, aber auch der Rechtsanwältinnen und Rechtsanwälte ein ganz wichtiger Faktor bei der Gestaltung dieser Prozesse ist. Ich habe ein hohes Vertrauen, dass es um einen vernünftigen, um einen an den verfassungsrechtlichen Vorgaben und den rechtlichen Rahmenbedingungen orientierten Einsatz geht. Entscheidend sind aber die Punkte, die in unterschiedlichen Stellungnahmen angeklungen sind, den Raum für die Betätigung dieser ethischen und rechtlichen Rahmenbedingungen zu schaffen, sodass wir überhaupt als Justiz eine Position gewinnen, die uns betreffenden Rahmenbedingungen aktiv mitgestalten zu können. Da werden wir ohne eine aktive Unterstützung von allen Seiten außerhalb auch der Justiz aus ganz unterschiedlichen Gründen nicht auskommen.

Das hat einmal den naheliegenden Grund, das wurde an vielen Stellen angesprochen, dass wir uns in die Entwicklung selber proaktiv einbringen müssen. Die Entwicklung wird Geld und Ressourcen kosten. Sie ist aus meiner Sicht aber unverzichtbar, damit wir all diese rechtlichen und ethischen Fragestellungen, die zu Recht thematisiert werden und die auch in den Fragen angelegt sind, mit sachgerechten Antworten versehen können.

Das andere Szenario sieht so aus, dass wir durch die Marktmacht von dem anderen Kontext entwickelten Lösungen irgendwann vor dem Dilemma stehen werden, dass wir sie allein zur Nutzung der Produktivitätsgewinne irgendwie in der Justiz implementieren müssen, ohne vorher aktiv mit in die Gestaltungsphase eingetreten zu sein.

Machen wir uns nichts vor: Wir sind auch in der Justiz in einem schwierigen Umgestaltungsprozess, was etwa die demografische Entwicklung angeht. Wir werden viele der Fragen der Arbeitsfähigkeit der Justiz und des Rechtsstaats insgesamt nicht lösen können, ohne dass wir im verstärkten Maße technische Prozesse einsetzen.

Jetzt haben wir noch die Chance, diese zu gestalten. Insofern möchte ich die erste Frage so beantworten: Ich denke, es ist für irgendwelche Regelungen noch zu früh. Wir sehen an der europäischen Diskussion über den AI Act, dass wir noch in der Findungsphase sind, was überhaupt ein angemessenes Regelwerk ist. Da sind wir meines Erachtens noch nicht so weit, dass wir in Nordrhein-Westfalen eigene Regeln setzen könnten, die dort einen prägenden Beitrag leisten könnten. Aber den Rahmen zu schaffen, dass sich Nordrhein-Westfalen in seinen Möglichkeiten innerhalb der Justiz einbringen kann, ist eine ganz wichtige Aufgabe, für die wir die Unterstützung mittel- und langfristig benötigen werden.

Ich komme damit gerne zu der zweiten Frage: „Warum ist ChatGPT für Urteile ungeeignet?“ – An die vielen Gründe, die wir gerade gehört haben, knüpfe ich an. Wenn Sie aber die Frage ein bisschen anders formulieren, bin ich mir nicht mehr ganz so sicher in meiner Antwort. Wenn die Frage lauten würde „Warum sind Large Language Models ungeeignet für Urteile?“, würde ich Ihnen antworten, ich glaube sie sind gar nicht so ungeeignet. Sie sind in der Entscheidungsfindung ungeeignet, da stimme ich mit allen Vorrednern überein, aus den besagten Gründen.

Ich möchte jetzt kurz auf den strafrechtlichen Bereich zurückkommen. Wir haben bisher viele zivilrechtliche Beispiele gehört. Der Punkt bis zur Urteilsfindung ist kein zeitliches, punktuelles Moment, sondern ein langer Prozess, der irgendwann bei einer Strafanzeige im Regelfall beginnt, über die polizeiliche Ermittlungsarbeit zur staatsanwaltlichen Betrachtung kommt, verbunden mit vielen Einschaltungen von Ermittlungsrichtern bis es am Ende zu einer richterlichen Entscheidung führt.

Meiner Überzeugung nach, und das zeigen auch die ersten praktischen Experimente, die wir in der ZAC NRW mit großen Sprachmodellen unternommen haben, sind auf dem Weg zum Urteil ganz viele Elemente machbar und einsetzbar, die mit großen Sprachmodellen gelöst werden können. Ich möchte es nur ganz kurz anreißen. Neben der Textgenerierung, die wir jetzt in den Blick genommen haben, kann ich mir durchaus im Bereich der Alltags- und Bagatellkriminalität mit überschaubaren Größenordnungen vorstellen, dass etwa eine Richterin oder ein Richter eine Entscheidung vorgibt und sie strukturiert zuliefert und am Ende nur noch die textliche Abfassung der getroffenen Entscheidung über ein Sprachmodell angereichert wird. Aber auf dem Weg dahin gibt es Dinge wie Beweismittlerschließung. Wir haben von den Fähigkeiten schon kurz gehört, dass Sprachmodelle sehr gut geeignet sind, um große Datenmengen textlicher Natur zusammenzufassen, zu abstrahieren und auch transparent insoweit mit Fundstellen zu belegen und damit eine bessere Suchfunktion zu ermöglichen. Das ist vielleicht in der technischen Wirkweise etwas verwirrend, trifft aber auf den Arbeitsalltag zu.

All diese Technologien können wir schon einsetzen, müssen uns aber von der Frage lösen, ob ich dem Computer sage, er solle mir ein Urteil schreiben. Meine Frage wäre zum jetzigen Zeitpunkt eher, welche Hilfsdienstleistungen mir ein großes Sprachmodell liefern kann.

Ich möchte einen letzten Punkt noch kurz ansprechen. Auch die Interaktion mit Rechtssuchenden hat durchaus Potenzial, zu ansprechender Gestalt zu werden, indem wir die „Kommunikationsfähigkeiten“ mit all den Grenzen, die das hat, dadurch optimieren, dass wir große Sprachmodelle einsetzen, um einfache Anliegen an die Justiz automatisiert beantworten zu können und damit schlicht einfach Rechtssuchenden schnelleren Zugang zu einfachen Möglichkeiten zu geben.

Aber, und damit komme ich zur nächsten Frage, welche Gefahren gehen damit einher? Natürlich ist das nicht gefahrenfrei. Über die Gefahren in der Urteilsfindung haben wir schon gesprochen. Da schließe ich mich den Vorrednern an. Wir haben aber noch eine ganze Reihe technischer und auch rechtlicher Rahmenbedingungen, die Gefahren mit sich bringen.

Zunächst muss man im Bereich der Betriebsorganisation einen geschärften Blick auf die Führungsverantwortung richten. Wir haben im staatsanwaltschaftlichen Bereich die Besonderheit, dass wir durchaus eine Weisungsgebundenheit haben, wenn auch mit gewissen Einschränkungen. Es ist natürlich Aufgabe von Führung, dafür zu sorgen, dass man die Effekte wie Automatisierungsgewöhnung, das blinde Vertrauen auf die Entscheidung des Computers, kennt, überprüft und Gegenmaßnahmen einleitet, so dass sie nicht im blinden Vertrauen auf technische Lösungen zu vermeintlichen Arbeitserleichterungen genutzt werden. Da braucht es einen intensiven Diskurs auch innerhalb

der Justiz selber, wie wir mit diesen Herausforderungen umgehen und wie wir diese Prozesse aktiv gestalten.

Große Schwierigkeiten sehe ich auch, und das muss ich einräumen, in den Betriebsmodellen. Wir haben mittlerweile eine Situation, wo wir jeden Monat eine komplett neue Generation von Sprachmodellen in der Veröffentlichung sehen, also eine sehr dynamische, eine sehr schnelle Entwicklung, von denen ein Großteil mit der Nutzung einer Public Cloud aus ressourcentechnischen Gründen, also letztlich mit dem Rechenzentrum eines Drittanbieters einhergeht. Dieses Betriebsmodell ist für die Justiz nicht uneingeschränkt nutzbar, weil wir Justizdaten, hoheitliche Daten nicht einfach Dritten anvertrauen können. Hier müssen wir intelligente Lösungen finden, um einerseits die vielen Potenziale, die das bietet, nutzen zu können, und uns nicht vom wissenschaftlichen Fortschritt abzukoppeln, andererseits aber die Datenhoheit weiter zu erhalten.

Der letzte Punkt in diesem Bereich: Wir haben auch große Herausforderungen, dass wir als Justiz hinreichend aktiv bleiben. Frau Biallaß hat es angesprochen. Wir haben ein unheimliches Privileg, dass wir einen Monopol-Datenbestand haben. Den hat kein anderer. Ohne uns wird niemand für die Justiz ein entsprechend taugliches Sprachmodell gestalten können. Wir müssen also sehr sorgfältig darauf aufpassen, dass wir das, was wir an Gewicht, an Gestaltungspotenzial haben, auch sachgerecht einsetzen und damit wirklich ein Gegengewicht gegenüber Weltkonzernen leisten können.

Einsatz von ChatGPT in anderen Ländern – Überlastung der Justiz. Herr Dr. Nink hat als Beispiel die Reichsbürger genannt. Nehmen wir mal die Vielschreibenden. Es ist durchaus ein bekanntes Phänomen. Im anwaltlichen Bereich, um das kurz einzuschleiben, mache ich mir weniger Sorgen. Da ist das Landesrecht relativ strikt, um Regelungen zu treffen, dass der rechtsmissbräuchliche Einsatz von solchen Technologien eingegrenzt werden kann. Die Möglichkeiten, die die anwaltliche Selbstverwaltung dazu bietet, sind ausreichend.

Natürlich haben wir auch Dritte, die an uns herantreten, Bürgerinnen und Bürger, die teils ihre Anliegen in hoher Vielzahl und ausführlicher Länge darbieten, die nicht immer sachgerecht und sinnvoll sind. Das bindet erhebliche Ressourcen bei der Justiz auf allen Ebenen. Aber hier ist meine Antwort ebenfalls: Natürlich können Rechtssuchende diese Technologie verwenden, um irgendwelche virtuellen Schriftsätze einzureichen und zu erstellen. Das hat es immer schon gegeben, aber wir haben als Justiz erstmalig die Möglichkeit, diese Technologien für den sachgerechten Einsatz gegen dieses rechtsmissbräuchliche Antragswesen einzubringen. Wir haben, um das aus der Praxis mitzubringen, mit einem Forschungspartner ein Modell gebaut, das es den Dezernentinnen und Dezernenten, die solche rechtsmissbräuchlichen Eingaben bearbeiten, ermöglicht, auch bei sehr vielseitigen Schriftsätzen auf dem Blick zu erkennen, was denn der neue Sachvortrag ist; denn ganz oft ist es repetitiv. Da werden Anliegen, die schon vor 20 Jahren das erste Mal angebracht wurden, wiederholt und noch mal wiederholt, ein bisschen sprachlich modelliert. Wir haben mit den Sprachmodellen und einigen anderen technologischen Entwicklungen nun technische Möglichkeiten, die wir als Justiz ganz aktiv nutzen können, um zu erkennen, was denn eigentlich an dieser Eingabe noch gelesen werden muss. Deswegen müssen wir meines Erachtens an der Stelle sagen: Jede Technologie hat Missbrauchspotenzial, das gilt für die großen

Sprachmodelle auch, aber unsere Aufgabe muss es sein, gegen Missbrauchspotenziale die Justiz zu ermächtigen, diese Technologien einzusetzen.

Sie hatten gefragt, ob die Justiz da aufrüsten muss. Ja, wir haben Möglichkeiten und Chancen, mit denen wir unser System und damit letztlich in gewisser Weise auch die Funktionsfähigkeit des Rechtsstaats sichern können.

In der Summe, glaube ich, müssen wir uns ein bisschen davon lösen, das ist allen Fragen gemein, nur den Blick auf ChatGPT zu richten. Das ist ein marktgängiges, die mediale Diskussion sicher beherrschendes Projekt. Die dahinterliegenden technologischen Chancen großer Sprachmodelle sind viel, vielgestaltiger als eben nur dieses eine Produkt. ChatGPT ist untauglich, die technologischen Lösungen können uns in vielen Dingen weiterbringen.

Prof. Dr. Simon J. Heetkamp (Technische Hochschule Köln, Professur für Wirtschaftsrecht, Mobilität und Versicherungsrecht [per Video zugeschaltet]): Auch von mir zunächst ein herzliches Dankeschön, dass ich die Stellungnahme zusammen mit dem Kollegen Schlicht abgeben konnte und bei der heutigen Sitzung dabei sein darf. Auch ich werde mich darauf beschränken, nur die Dinge hinzuzufügen, die von der Kollegin und den Kollegen noch nicht genannt wurden. Die erste Frage ging dahin, Herr Dr. Pfeil, wie die Legislative den einzelnen vor der falschen Anwendung der KI durch die Richter schützen könnte. Ich glaube, Sie haben Recht, wenn Sie sagen, wir brauchen einen gesamtgesellschaftlichen Dialog, und die Kollegin und die Kollegen haben Recht, wenn sie sagen, die richterliche Unabhängigkeit und das Berufsethos sind hochzuhalten.

Ich würde auch denken, dass der einzelne im einzelnen Verfahren durch den Instanzenzug und die Grundrechte gut genug geschützt ist, und es keines weiteren Tätigwerdens durch die Legislative braucht. Allerdings könnte man darüber nachdenken, ob man nicht eine Art Code of Conduct, also einen Verhaltenskodex für die Richter entwickelt. Das sieht man jetzt schon teilweise im anwaltlichen Bereich, wo sich entweder einzelne Kanzleien oder auch Anwaltsvertretungen, Berufsvertretungen von Anwältinnen und Anwälten dazu entschließen, entsprechende Verhaltenskodexe zu formulieren. Wer das jetzt für den richterlichen Bereich machen sollte, ob das auch eine richterliche Berufsvertretung oder das Justizministerium oder Ähnliches wäre, wäre dann die zweite Frage.

Ich fand den Hinweis der Kollegin Biallaß sehr gut, dass wir, wenn es zu dem Einsatz von großen Sprachmodellen der Justiz kommt, dann umfangreiche Schulungen für die Richterinnen und Richter haben müssten, was natürlich auch die entsprechenden Angebote und finanziellen Mittel voraussetzt. Wir haben es bei der Einführung der E-Akte gesehen, dass die Schaffung der Akzeptanz bei den Richterinnen und Richtern ein ganz wesentlicher Faktor ist. In diese Richtung geht auch der Hinweis von Herrn Grabmair, der sagt, es müssen nicht alle Leute das sogenannten Prompten lernen, wie ich ein großes Sprachmodell bediene, sondern vielleicht würde man bei der Entwicklung eines großen Sprachmodells für die Justiz genau das mitdenken, dass man spezielle Funktionen entwickelt und dann einfach per Schaltfläche hinterlegt, die dann alle bedienen könnten.

Das bringt mich schon zur zweiten Frage, warum ChatGPT zur Entscheidungsfindung – ich erweitere hier die Frage – oder Entscheidungsunterstützung geeignet oder nicht geeignet ist. Das würde ich gerne, weil die Frage für mich ein wenig zu negativ formuliert war, ins Positive wenden, nämlich dass – Herr Hartmann und Herr Grabmair haben schon darauf hingewiesen – große Sprachmodelle erheblichen Mehrwert bieten, und zwar selbst dann nicht, wenn man die Urteile nicht damit schreibt. Man kann vielmehr gar nicht genug betonen, was es für einen Mehrwert hätte, wenn ich von einer großen Bauakte mit 500 oder 600 Seiten mit einem Klick eine Chronologie erstellen lassen könnte, die ich nur noch überprüfen muss, die ich beim Lesen der Akte daneben lege und nur noch überprüfe, ob das Programm die richtigen Daten herausgesucht hat. Wenn ich das händisch mache, brauche ich wahrscheinlich einen Arbeitstag dafür. Wenn ich jedoch beim Lesen nur noch überprüfe, ist es viel weniger Arbeit. Es könnte ein richtiger Gamechanger sein.

Herr Grabmair hat in seiner Stellungnahme einen sogenannten Frage-Antwort-Assistenten angesprochen. Wenn man eine Akte liest, das kann ich aus der richterlichen Tätigkeit berichten, ist es häufig so, dass man weiß, es irgendwo schon gelesen zu haben. Wenn man dann einen Assistenten zum Beispiel fragen könnte, wo der Kläger zu dem und dem Thema ausführt, dann bin ich direkt da. Das wäre eine viel, viel bessere Suche. Auch hier würden wir das Fundament, das wir mit der E-Akte gelegt haben, tatsächlich bebauen und nutzen, weil viele Kolleginnen und Kollegen manchmal fragen, was denn die E-Akte bis jetzt für einen Mehrwert biete. Das wäre ein ganz erheblicher.

Im Rahmen dieser Frage würde ich auch gerne darauf hinweisen, dass bis jetzt alle Fragen und viele Fragen des Fragenkatalogs immer auf den richterlichen Bereich abzielen. Ich würde mir wünschen, dass wir auch den nichtrichterlichen Bereich bedenken. Auch da gibt es meines Erachtens ein großes Potenzial für große Sprachmodelle, wenn man einfach nur an die Veraktung oder an das Auslesen und Verfassen von Metadaten, was bisher sehr zeitaufwendig, sehr händisch und sehr monoton durch die Serviceeinheiten erfolgt, denkt.

Die dritte Frage war, ob schon heute Sprachtools in der Justiz eingesetzt werden und wenn ja, welche. – Dazu hat die Kollegin Biallaß schon ausgeführt und einige Modelle oder KI-Tools benannt. Wir hatten dazu im Januar schon eine Sachverständigenanhörung, wo die entsprechenden Tools ausführlich beschrieben wurden. Hierzu würde ich gerne ergänzen, dass wir im Rahmen unserer digitalen Richterschaft – also diese Austauschplattform für Digitalisierungssysteme für Leute, die in der Justiz beschäftigt sind – eine Umfrage unter den Kolleginnen und Kollegen gestartet haben. Über 600 Leute waren dabei und wurden gefragt, welche Erfahrungen die Kolleginnen und Kollegen bereits mit ChatGPT, ChatGPT-4 oder einfach mit großen Sprachmodellen gemacht haben.

Viele Kolleginnen und Kollegen haben zurückgemeldet, dass sie es mal spielerisch mit anonymisierten Sachverhalten und Rechtsfragen ausprobiert haben und dann doch die Ernüchterung relativ groß war, dass das Programm in der Tiefe in der Domäne des deutschen Rechts noch nicht zu gebrauchen war. Aber Kollegen aus Baden-Württemberg haben darauf hingewiesen, dass sie ein anderes Tool nutzen, was ich auch in der

Stellungnahme beschrieben haben, nämlich DeepL Write, was einem hilft, Sachen, die man selbst geschrieben hat, umzuformulieren und zu verbessern, sodass die sprachliche Verständlichkeit für die Leserin, den Leser verbessert wird und der rechtsuchende Bürger, oder die Bürgerin es besser versteht. Das ist vielleicht nur ein kleines Beispiel, was nicht groß etwas ändern wird, aber es hat mir gezeigt, dass viele Kolleginnen und Kollegen auf der Suche nach technologischer Unterstützung in der täglichen Arbeit sind und das, was wir hier besprechen, wahrscheinlich auf sehr offene Ohren in der Richterschaft stoßen wird.

Die letzte Frage war, ob ChatGPT oder große Sprachmodelle die Justiz blockieren würden. Ich wäre hier entspannt. Das sehe ich erst einmal als nicht so kritisch an. Der Kollege Grabmair hat zu Recht darauf hingewiesen, dass zum einen ChatGPT-2.0 das vielleicht schon konnte, aber mir steht vielmehr meine Erfahrung als Rechtsanwalt vor Augen. Ich war in dem sogenannten Dieserverfahren tätig, und da gab es auch Programme mit einem Frage-Antwort-Katalog. Man musste ein Dutzend Fragen beantworten und hat direkt einen Schriftsatz bekommen, der hundert Seiten lang war. Mit wenig Arbeit, drei Minuten Arbeit, hat man hundert Seiten Schriftsatz, die man dann verschicken konnte. Das ist jetzt durch große Sprachmodelle ein anderer Ansatzpunkt, wird aber die Justiz in Gänze nicht überfordern. Wenn das zu einer großen Anzahl von Verfahren zusätzlicher Art führen würde oder die Schriftsätze deutlich länger würden, könnte man mit entsprechender Personalausstattung reagieren.

Zu dem Einzelfall, der jetzt auch in den Medien war, mit diesen halluzinierenden Präzedenzfällen wurde auch schon etwas gesagt, dass es im deutschen Recht nicht dieselbe Relevanz hat. Dazu möchte ich aus der Erfahrung, die ich als Richter in den Verfahren zu Fluggastrechten gemacht habe, ergänzen. Es ist häufig so, dass die Klägerkanzleien, aber auch die Beklagtenseite natürlich Spezialisten sind und in vielen Hunderten oder Tausenden Verfahren aktiv sind und häufig über sehr gute Urteilsdatenbanken verfügen, viel besser als das einzelne Gericht es hat. In den entsprechenden Schriftsätzen werden häufig Urteile von anderen Gerichten zitiert, die weder über öffentliche Datenbanken zugänglich sind noch über „weg-online“ oder „juris“ und dem Gericht auch nicht vorliegen. In solchen Fällen kann ich einfach die entsprechenden Urteile, die der Kläger oder der Beklagte anführen, ignorieren oder erlasse einen entsprechenden Hinweis, dass sie das Urteil vorlegen sollen, wenn es ihnen darauf ankommt. Dann kann man überprüfen, ob das Urteil eine Relevanz hat oder nicht. Ansonsten wäre es so, wie die Kollegin Biallaß gesagt hat, dass häufig auch Urteile angeführt werden, die, wenn man sie in Gänze betrachtet, auch für den zu entscheidenden Sachverhalt gar nicht so relevant sind.

Der letzte Punkt, den ich an der Stelle nennen möchte, ist die Relevanz, die große Sprachmodelle vielleicht auch für die Bürger haben könnten, nämlich in den Bereichen, in denen wir keinen Anwaltszwang haben, wo Bürger eigenständig tätig werden können. Ich würde da gar nicht so sehr das Augenmerk auf Reichsbürger oder andere Querulanten legen, sondern vielmehr auf den Bürger, der sich den Anwalt sparen will. Hier müssen wir uns die Frage stellen, wie sich das auf das Zivilverfahren auswirken wird, wenn eine höhere Anzahl von Bürgerinnen und Bürger versucht, ihr Recht durchzusetzen und durch die großen Sprachmodelle schlecht beraten werden.

Vorsitzender Dr. Werner Pfeil: Herr Professor Heetkamp, vielen Dank. Damit haben wir die erste Fragerunde abgeschlossen und die zweite Runde würde sich anschließen. Ich möchte darauf hinweisen, dass wir die Anhörung um 15 Uhr beenden. Es sollten nur noch Einzelfragen an die Sachverständigen gestellt werden, keine Fragen mehr an alle gemeinschaftlich, da es ansonsten mit der Zeit nicht funktionieren wird. Die CDU-Fraktion darf beginnen.

Angela Erwin (CDU): Ich habe gar keine Frage mehr, sondern möchte nur eine kleine Anmerkung machen. Wir haben gerade bei ChatGPT die Frage eingegeben: Kann man mit ChatGPT Rechtsurteile verfassen? – Die eigene Antwort ist: Als KI-Sprachmodell kann ich Texte generieren einschließlich rechtsbezogener Inhalte, allerdings bin ich kein Anwalt und meine Antworten sollten nicht als rechtsverbindliche Ratschläge oder offizielle Rechtsurteile betrachtet werden. Zur rechtsverbindlichen Information und Rechtsberatung sollten Sie sich an einen qualifizierten Rechtsanwalt oder eine Rechtsanwältin wenden.

Das wollte ich der Runde hier nicht vorenthalten.

Sonja Bongers (SPD): Noch einmal recht herzlichen Dank für die weiteren Informationen. Wir haben keine Fragen mehr.

Dagmar Hanses (GRÜNE): Ich habe noch eine Frage an Herrn Hartmann. Er hat insbesondere den Gestaltungszeitraum angesprochen, dass wir jetzt in einem spannenden zeitlichen Gap sind, in dem wir jetzt noch Dinge gestalten können. Könnten Sie das noch näher ausführen, was Sie sich erhoffen, was jetzt zu gestalten wäre?

An Herrn Professor Grabmair und an Frau Biallaß habe ich die Frage nach der Fortbildung. Ich denke, wir sind uns alle einig, dass hier ein Fortbildungsbedarf besteht. Wir haben die Justizakademie. Die Fortbildungspflicht für Richterinnen und Richter ist im Richter- und Staatsanwältegesetz für das Land NRW verankert. Wie müssen die Angebote aus Ihrer Praxis aussehen, damit sie insbesondere bei Richterinnen und Richtern ankommen?

Vorsitzender Dr. Werner Pfeil: Für die FDP-Fraktion schließe ich mich Frau Hanses an. Herr Hartmann, Sie haben zu dem Rahmen und der Gestaltungsphase, insbesondere zu dem Rahmen, den NRW sich selber geben sollte, etwas gesagt. Ich bitte Sie, dies näher zu konkretisieren.

An Herrn Nink und Frau Biallaß habe ich noch folgende Frage: Sehen Sie die KI-Verordnung, die derzeit im Trilog ist, kritisch?

Prof. Dr. Daniel Zerbin (AfD): Wir haben auch noch eine Frage. Es ist ein interessantes Thema, wir beschäftigen uns im Wissenschaftsausschuss am kommenden Montag auch mit ChatGPT und KI. Ich habe eine Frage, was die Akzeptanz und Belastung angeht. Wir haben gerade von Herrn Professor Heetkamp gehört, dass es im Justizsystem eine Innovation auch bei Rechtsanwälten gibt. Da treffen zwei Systeme aufeinander.

Auf der einen Seite habe ich Rechtsanwälte, die auf dem freien Markt sind, die sich anpassen müssen, und auf der anderen Seite habe ich das System Verwaltung und die Justiz. KI arbeitet von der Progression her nicht linear, sondern exponentiell. Da ist die Frage an Frau Biallaß und Herrn Hartmann: Sehen Sie eine gute Möglichkeit, dass sich die Justiz anpassen kann, dass sie nicht überstrapaziert wird?

Frau Biallaß, Sie hatten von Digital Competences gesprochen. Wie wird die Akzeptanz von solchen Systemen in Zukunft sein? Würde es auch Probleme geben? Wir haben gerade über die E-Akte gesprochen. Wird sich die Justiz entsprechend anpassen können und auch akzeptieren, andere Systeme einzuführen?

Vorsitzender Dr. Werner Pfeil: Vielen Dank. – Damit gehen wir in die zweite Antwortrunde und beginnen mit Herrn Dr. Nink.

Dr. David Nink: An mich gerichtet war, wenn ich es richtig sehe, nur Ihre Frage, Herr Dr. Pfeil zur EU-KI-Verordnung. Mit Blick auf die fortgeschrittene Zeit möchte ich mich gerne kurzfassen. Im Grundsatz ist es sehr begrüßenswert, auf europäischer Ebene in die Initiative zu gehen. Inhaltlich sehe ich es bislang noch ein bisschen kritisch, zum einen, weil sich die Definitionen noch ändern und noch nicht so ganz trennscharf sind. Speziell die Definition des Kerns des Gegenstandes KI ist sehr weit. Es gibt diese Einordnung in Hoch-Risiko-KI. Ab dann wird es ein bisschen abgestuft, und man hat ein gestuftes Regelungsregime von Verboten bis hin zu Transparenzpflichten. Nach jetzigem Stand ist es so, dass sehr viele auch jetzt schon eingesetzte Softwaresysteme unter dem KI-Begriff fielen und dementsprechend auch sehr hohe Anforderungen mit sich brächten. Das sehe ich kritisch auch für mittelständische Unternehmen in Deutschland, da Schritt zu halten. Aber im Grundsatz ist die Regelungsinitiative durchaus begrüßenswert.

Isabelle Biallaß (Deutscher EDV-Gerichtstag): Ich beginne mit der Frage zur Fortbildung. Zum einen möchte ich an das anschließen, was Herr Professor Heetkamp gerade gesagt hat und ich in meiner schriftlichen Stellungnahme thematisiert hatte. Ich halte es für sehr hilfreich, wenn man allen Justizmitarbeitern eine Art Leitfaden an die Hand gibt, wie es schon viele private Unternehmen tun, mit dem man dafür sensibilisiert, zum jetzigen Zeitpunkt bitte nicht privat irgendwelche Arbeitsdaten bei ChatGPT einzugeben und das als Bausteine für die Arbeit zu verwenden.

Mittelfristig halte ich es für notwendig, dass wir tatsächlich in unserer JAK strukturierte Fortbildungen aufbauen, wo man erst einmal für interessierte Kollegen Informationen auf freiwilliger Basis zur Verfügung stellen kann. Wenn wir irgendwann ein eigenes Large Language Model für die Justiz hätten, dann wäre meine Vorstellung, dass man den Roll-out so beginnt, wie wir es jetzt auch schon bei der E-Akte haben oder damals bei der Einführung von JUDICA und TSJ gemacht haben, dass es, bevor man das Programm auf seinem Rechner hat, eine Fortbildung gibt. Diese Fortbildung müsste im Gegensatz zu unseren bisherigen klassischen IT-Fortbildungen meiner Ansicht nach tatsächlich interdisziplinär vorbereitet werden. Wir bräuchten dafür Psychologen, die uns helfen, aufzuklären, wie wir die Mitarbeiter an der Stelle tatsächlich sensibili-

sieren und auf Gefahren hinweisen. Das muss man zugleich bei der Entwicklung eines entsprechenden Large Language Models für die Justiz machen, dass man bei der Gestaltung der Benutzeroberfläche, das, was Herr Professor Grabmair schon über CoCounsel geschildert hat, abbildet, dass man Warnhinweise verfasst, gerade wenn wir die Möglichkeit geben würden, selber zu prompten, dass die Ergebnisse völlig unzutreffend und halluziniert sein können, dass man sie für bestimmte Zwecke nicht verwenden kann oder Ähnliches. Es wird sehr anspruchsvoll, entsprechende Fortbildungskonzepte zu entwickeln.

Dann komme ich zu dem Punkt der KI-Verordnung. Da würde ich mich angesichts der fortgeschrittenen Zeit vollumfänglich den Ausführungen von Herrn Dr. Nink anschließen. Grundsätzlich halte ich es für gut und richtig, dass man auf EU-Ebene versucht, eine Lösung zu finden. Es gibt in allem, was bis jetzt veröffentlicht ist, Punkte, die ich anders sehe, insbesondere die Frage zwischen einerseits Regelungen treffen und andererseits aber Innovationen ermöglichen, ist ein sehr schmaler Grat. Dazu könnte man eine komplett eigene Anhörung machen.

(Zuruf von Dagmar Hanses [GRÜNE])

Vorsitzender Dr. Werner Pfeil: Frau Hanses, dafür sind wir nicht zuständig. – Frau Biallaß.

Isabelle Biallaß (Deutscher EDV-Gerichtstag): Das sollte keine Steilvorlage sein.

Dann kommen wir zu dem Punkt, wie man mit den Möglichkeiten derartiger Large Language Models umgeht. Würden sie von der Justiz angenommen, wenn man ein entsprechendes Modell anbindet? – Ich kann schon jetzt sagen, dass ich in meiner dienstlichen Eigenschaft Anfragen bekommen habe, ob wir nicht in die Gerichte gehen und erklären können, wie man ChatGPT in der Justiz einsetzen kann, und das kam nicht von jungen Kollegen. Es besteht angesichts des Medienechos ein Interesse daran, zu lernen, ob man es nicht sicher einsetzen kann.

Wenn wir eine sichere Alternative finden würden, hätten wir sicher einen großen Teil, der das gerne nutzen möchte. Es gibt auch welche, die sagen, sie möchten weiter klassisch wie bisher arbeiten, aber ich glaube, die Gruppe ist größer als es sich viele vorstellen können. Wir Mitarbeitende in der Justiz sind gar nicht so unaufgeschlossen für derartige Dinge, wie manche es uns gelegentlich unterstellen wollen. Dementsprechend sehe ich hier massive Chancen für die Einsatzmöglichkeiten, wenn wir eine vernünftige für die Justiz passende Lösung entwickeln können.

Prof. Dr. Matthias Grabmair (Technische Universität München, TUINF27 Informatik 19 – Professur für Legal Tech [per Video zugeschaltet]): Ich kann mich vollumfänglich Frau Biallaß mit ihren Ausführungen zur Fortbildung anschließen. Ich glaube, das war die einzige Frage, die auch an mich gerichtet war. Ich kann dem nur noch hinzufügen, dass es in der Natur der technischen Entwicklung liegt, dass sie voranschreitet, und sich die Möglichkeiten weiterentwickeln werden. An der Stelle möchte ich hinzufügen, dass sofern diese Fortbildungsinhalte generiert werden, sie idealerweise

in ein ganzheitliches Konzept gebettet würden, wonach eine längerfristige Beobachtung und ein Monitoring möglich wären und die Rechtspflege den Anwendern des Modells regelmäßig Feedback gibt, was gut funktioniert hat und was nicht gut funktioniert hat, welche neuen Abkürzungen und Arbeitsabläufe sich eingeschlichen haben usw. Eine Fortbildung muss auch ein entsprechendes Monitoring beinhalten. Ein besseres deutsches Wort fällt mir dafür leider nicht ein. Man muss die Sache beobachten und zyklisch neu bewerten und anpassen, sodass es längerfristig nachhaltige positive Entwicklungen gibt.

Markus Hartmann (Zentral- und Ansprechstelle Cybercrime Nordrhein-Westfalen):

Ich beginne mit der letzten Frage, weil ich die im Anschluss an die Ausführungen von Frau Biallaß beantworten kann. Ich glaube auch, dass wir in der Justiz gerade ein befindliches Mitarbeiterfeld auf der Suche haben, die solche Technologien dankbar aufnehmen werden. Es wird Herausforderungen geben, aber das Anpassungsvermögen der Justiz in solche Technologien halte ich tatsächlich auch für groß und sehe da nicht das Risiko einer Überforderung.

Was die Frage des Gestaltungsprozesses angeht, möchte ich mit dem beginnen, was Sie völlig zu Recht sagen, Frau Hanses. Wir haben jetzt das Zeitfenster, und das ist der Punkt, der in den Vordergrund meiner Betrachtung rücken soll. Wir haben jetzt ein Zeitfenster, wo wir eine sehr dynamische, wirtschaftliche und wissenschaftliche Entwicklung haben und wir gleichzeitig ein sehr interessantes Gegenüber sind. Das merken wir in dem Dialog, den wir mit Unternehmen und wissenschaftlichen Akteuren führen, dass wir ein sehr spannendes Gegenüber sind, weil wir einen extrem spannenden Datensatz haben, der zur Erforschung und Entwicklung ansteht.

Um das Skizzieren zu wollen, brauchen wir im Grunde drei verschiedene Dinge, die wir angehen müssen. Der erste Punkt ist, dass wir möglichst schnell eine interne Diskussion führen, wohin wir als Justiz wollen, dass wir eine interne fundierte Auffassung bilden, bei welchen Punkten wir uns priorisiert insgesamt als Justiz den Einsatz großer Sprachmodelle vorstellen können, und welche Punkte schnell und einfach umsetzbar sind, die Quick-Wins, und wohin mittel- und langfristig die Entwicklung gehen muss.

Nach dieser Meinungsbildung brauchen wir einen zweigestalteten Prozess. Der eine Prozess ist tatsächlich, dass wir ins Tun kommen müssen, weil wir das Gestalten nur durch das Mitarbeiten erlernen werden. Ich denke, wir brauchen ein dichtes Netzwerk von Wirtschafts- und von Wissenschaftspartnern von beiden Seiten, die an kleinen handhabbaren Projekten genau diese technologische Umsetzarbeit liefern. Das ist einmal wichtig, um die Erfahrung damit zu gewinnen, was das im praktischen Arbeiten heißt. Wir reden jetzt auf einer theoretischen Ebene. Wir brauchen die praktischen Einsichtnahmen. Es gibt uns auch die Möglichkeit, hinsichtlich der Frage nach der Akzeptanz in der Justiz, einfache, schnelle Lösungen für konkrete abgegrenzte Probleme zu bieten und diese technologisch umzusetzen. Das heißt auch, Akzeptanz bei Kolleginnen und Kollegen zu steigern, weil Technologie nicht abstrakt, sondern erfahrbar wird. Das ist sozusagen die zweite Ebene: tun und machen.

Bei der dritten Ebene bin ich fest davon überzeugt, dass wir eine Art übergreifenden Dialog brauchen. Wir müssen als nordrhein-westfälische Justiz gegenüber Wissenschaft

und Wirtschaft als Akteur in Erscheinung treten. Als solcher werden wir noch nicht uneingeschränkt wahrgenommen. Wir müssen also, etwas plakativ gesprochen, die Arme weit ausbreiten und sagen: Wir wollen uns in dem Bereich engagieren, wir wollen als Justiz dort eine namhafte Rolle spielen. Dann, denke ich, werden wir eine kritische Masse aus der unternehmerischen Seite für diese Technologie mitbringen, aber auch aus der wissenschaftlichen Begleitung haben, die gemeinsam mit unserem Fachwissen mehr erreichen können als jeder für sich alleine könnte. Denn wir werden als Justiz nicht alleine dieses Thema retten und bewegen können, aber umgekehrt müssen wir uns auch bewusst sein, dass auch die Wirtschaft und Wissenschaft alleine ohne uns dieses Thema mit Justizkontext nicht bewegen wird. Wenn wir diese drei Elemente – interner Diskurs, wo wollen wir hin, konkretes Machen und Tun im kleinen Rahmen zur Erfahrungsgenese und einen übergreifenden Dialog – in diesem Bereich zusammenbringen und ihnen in Nordrhein-Westfalen sozusagen eine programmatische Heimat geben, dann bin ich recht davon überzeugt, dass es gerne angenommen wird.

Vorsitzender Dr. Werner Pfeil: Herr Hartmann, vielen Dank. Das waren alle Fragen und die darauf gegebenen Antworten. Weitere Fragen sehe ich nicht.

Ich bedanke mich bei den Sachverständigen, die zugeschaltet waren, und bei denen, die hier vor Ort waren zum einen für die schriftlichen Ausführungen und zum Zweiten, um heute hier die weiteren Fragen beantworten zu können. Ich bedanke mich bei den Abgeordneten für Ihre Anwesenheit heute hier. So warm ist es gar nicht mehr, die nächsten drei Plenartage werden sicherlich wärmer. Ich wünsche allen einen schönen Nachmittag und eine gute Heimfahrt.

gez. Dr. Werner Pfeil
Vorsitzender

Anlage

10.08.2023/10.08.2023

Anhörung von Sachverständigen
des Rechtsausschusses**Einsatz von ChatGPT im Justizbereich**
Vorlage 18/1022am Dienstag, dem 13. Juni 2023
13.30 bis (max.) 15.00 Uhr, Raum E3 A02, Livestream**Tableau**

eingeladen	Teilnehmer/innen	Stellungnahme
Professor Dr. Christoph Johannisbauer Hochschule für Polizei und öffentliche Verwaltung Nordrhein-Westfalen Abteilung Duisburg Duisburg	Prof. Dr. Christoph Johannisbauer <i>(per Videozuschaltung)</i>	18/590
Dr. David Nink Hundsangen	Dr. David Nink	18/592
Isabelle Désirée Biallaß Vorstandsmitglied EDV Gerichtstag Richterin am Amtsgericht Essen Lehrbeauftragte FH Köln <i>Essen</i>	Isabelle Biallaß	18/575
Professor Matthias Grabmair Professur für Legal Tech TUINF27 Informatik 19 - Professur für Le- gal Tech (Prof. Grabmair) Garching b. München	Prof. Dr. Matthias Grabmair <i>(per Videozuschaltung)</i>	18/588
Markus Hartmann Leitender Oberstaatsanwalt Leiter der ZAC NRW Generalstaatsanwaltschaft Köln Zentral- und Ansprechstelle Cybercrime Nordrhein-Westfalen - ZAC NRW - Köln	Markus Hartmann	18/571

eingeladen	Teilnehmer/innen	Stellungnahme
Professor Dr. Simon J. Heetkamp, LL.M. Richter am Landgericht Mediator (Hochschule Darmstadt) Professur für Wirtschaftsrecht, Mobilität- und Versicherungsrecht Technische Hochschule Köln Campus Südstadt Köln	Prof. Dr. Simon Heetkamp <i>(per Videozuschaltung)</i>	18/574
Dr. Christian Schlicht Köln	<i>keine Teilnahme</i>	