

**„Ich glaube, dass der richtige Umgang mit künstlicher Intelligenz auch in der Schule unterrichtet werden sollte.“ (Dr. Slav Petrov)**

***Benjamin Schmorl:** Ihre Familie stammt aus Bulgarien. Wann sind Sie nach Deutschland gekommen, und wann wurden Sie in Berlin eingeschult?*

**Slav Petrov:** Ich bin kurz nach der Wende, im Sommer 1991, im Alter von zehn Jahren nach Deutschland gekommen. Mein Vater kam schon ein halbes Jahr vorher nach Berlin. Wir hatten aber Probleme mit dem Visum und kamen daher erst später. Ich wurde in die 4. Klasse der Alt-Schmargendorf-Grundschule eingeschult. Ich hatte diese Klasse schon in Bulgarien absolviert, konnte aber so gut wie kein Deutsch und habe sie daher wiederholt. Als Vorbereitung hatte ich zwar von Frühling bis Sommer 1991 ein- bis zweimal pro Woche privaten Deutschunterricht – richtig viel konnte ich danach aber noch nicht. Ich vermute, das war von meinem Alter her die Grenze, diesseits derer ich Deutsch noch relativ akzentfrei lernen konnte.

***B.S.:** Beginnen wir unser Gespräch mit einem kurzen Rückblick auf Ihre Schulzeit am Arndt-Gymnasium Dahlem. Warum wurde gerade diese Schule ausgewählt, und welche Erinnerungen an einzelne Lehrkräfte und besondere Ereignisse kommen Ihnen spontan in den Sinn? Welche Leistungskurse hatten Sie in der Oberstufe bis zum Abitur im Jahr 2000?*

**S.P.:** Im Ostblock wurde ja traditionell großer Wert auf eine bestmögliche Ausbildung gelegt. Meine Eltern haben sich deshalb nach unserer Ankunft in Berlin erkundigt, welche Schulen die besten wären. Zuerst hatten wir vom Grauen Kloster gehört, das den Ruf einer sehr guten Schule hatte. Meine Eltern waren jedoch besorgt, weil hier gleich mit Latein und Altgriechisch begonnen wird. Sie wollten Englisch als erste Fremdsprache. Auf das Arndt-Gymnasium sind sie dann wegen des guten Rufs gekommen, da es einerseits ein humanistisches Gymnasium und andererseits auch in den Naturwissenschaften stark ist. Für meine Eltern, beide Chemiker, war das wichtig. Außerdem ist das AGD in Dahlem nicht sehr weit von Schmargendorf entfernt. Ich war dann auf dieser Schule sehr zufrieden, und mein Bruder Anton ist mir später ans Arndt-Gymnasium gefolgt und hat dort 2006, sechs Jahre nach mir, sein Abitur abgelegt.

Ich fand die Größe der Schule schön – man kannte sich in der Schülerschaft und nach einer gewissen Zeit kannte man auch viele Lehrer. Mich haben im Unterricht die Naturwissenschaften am meisten interessiert. Deshalb weisen meine Zeugnisse in der linken und der rechten Spalte meist sehr unterschiedliche Noten auf. Am AGD habe ich die 9. Klasse übersprungen, daher kannte ich dann zwei Jahrgänge. Insgesamt habe ich viele gute Erinnerungen an meine Mitschüler, und als später mein Bruder Anton dazukam, lernte ich auch viele seiner Freunde kennen, was sich sehr familiär angefühlt hat. Als ich dann in

der 11. Klasse als Begleiter an der traditionellen Skireise der 8. Klassen teilgenommen habe, hat sich mein Bekanntenkreis nochmals erweitert.

Meine Leistungskurse waren Mathematik und Physik, und ich habe sehr gute Erinnerungen an Herrn Meyer, der vor der Oberstufe auch mein Klassenlehrer war und letztlich viel Einfluss auf meine spätere Laufbahn hatte. Als ich in der 12. Klasse war, hat er mir einen Zeitungsschnipsel mitgebracht, auf dem stand, dass es für Schüler möglich sei, an der FU Informatikkurse zu belegen. Das habe ich dann auch gemacht. Also habe ich schon während der Schulzeit Grundvorlesungen in Informatik besucht, was mein nachfolgendes Studium beschleunigt und dazu beigetragen hat, dass ich ein Stipendium für die USA bekommen habe. Ohne Herrn Meyer wäre das nicht geschehen. Ein anderer Lehrer, an den ich viele gute Erinnerungen habe, ist Herr Speck-Hempel. Noch heute erinnere ich mich an viele Physik-Experimente in seinem Unterricht. Es gab einmal eine Klausur, in der es darum ging, weshalb man Fett in eine Pfanne gibt, wenn man ein Steak anbrät. Ich meinte: „Damit es nicht kleben bleibt!“ Es ging jedoch eigentlich um Thermodynamik, also darum, dass Fett besser Wärme leiten kann als Luft. Seitdem habe ich nicht vergessen, dass man auch in einer Teflon-Pfanne etwas Fett braucht!

***B.S.:** Ist in dieser Zeit bereits Ihr Interesse an Computern und künstlicher Intelligenz entstanden? Gab es dabei Schlüsselerlebnisse? Welche Rolle hat der Fußball gespielt?*

**S.P.:** Ja, Schlüsselerlebnisse gab es auf jeden Fall. Eines davon war folgendes: Ich hatte schon immer sehr viel Interesse an Mathematik und habe einmal in den Herbstferien an der FU an einem Seminar im Mathe-Fachbereich teilgenommen. Wir haben dort wochenlang einen Beweis zu führen versucht, dass man jede Landkarte mit vier Farben so ausmalen kann, dass zwei benachbarte Länder stets verschiedene Farben haben. Das ist ein wichtiges Resultat, aber der mathematische Beweis ist sehr schwer. Ich fand es deutlich spannender, einen Computer-Algorithmus zu entwickeln, mit dem man den Versuch dann auch wirklich umsetzen kann. Der Beweis selbst wurde mir irgendwann langweilig, und deshalb habe ich mir gesagt: „OK, ich mag Mathematik, aber ich mag mehr die angewandte Mathematik und nicht nur Beweise als Selbstzweck.“ Für mich wurde so relativ schnell klar, dass ich Informatik studieren möchte.

Die Verbindung zu Fußball und künstlicher Intelligenz kam etwas später, und zwar folgendermaßen: Im ersten Jahr an der Uni habe ich zufällig in einem Informatikgebäude von Studierenden programmierte Fußballroboter entdeckt, die sich auf Kunstrasen autonom bewegten. Das fand ich sehr spannend und habe gefragt, ob ich mitmachen könnte. Zuerst wurde dies mit dem Hinweis abgelehnt, dass ich erst einmal mein Vordiplom machen solle. Ich bin jedoch hartnäckig geblieben und war dann einige Male dabei. Dies war im Jahr 2001, zu Beginn meines Studiums, meine erste Erfahrung mit KI. 2002/2003 war ich dann für ein Jahr in den USA und weiß noch, dass in der Bewerbung für das

damit verbundene Stipendium Roboterfußball das Thema war. Später habe ich auch meine Diplomarbeit in diesem Bereich verfasst und 2004 das Studium beendet. Auf meiner Diplomurkunde ist übrigens ein Stückchen ausgeschnittener Rasen von dem Fußball-Feld angeheftet, auf dem 2004 in Portugal der sog. Robo-Cup, eine damals jährlich ausgetragene Weltmeisterschaft im Roboterfußball, ausgetragen wurde. Bei dieser Weltmeisterschaft gab es verschiedene Ligen. Wir hatten zwei Mannschaften in diesen Ligen: die „FU Fighters“. Die eine hat auf einem Feld gespielt, das ungefähr die Größe einer Tischtennisplatte hatte. Es gab eine Kamera von oben, die die Roboter beobachtete und einen zentralen Computer, der alle Roboter steuerte. Hier sind wir Weltmeister geworden! Dann gab es noch eine Liga mit größeren Plätzen – ungefähr so groß wie ein Tennisplatz. Die Roboter hatten eine Größe von ca. 60 Zentimetern. Sie hatten jeweils eine eigene Kamera und einen eigenen Laptop und mussten dann alleine miteinander kommunizieren. Hier habe ich ebenfalls mitgemacht, und wir sind bis ins Halbfinale gekommen.

***B.S.:** Kehren wir noch einmal zu Ihrer Studienzeit zurück. Welche Fächer haben Sie nach dem Abitur belegt? An welchen Universitäten? Wie haben Sie Ihr Studium abgeschlossen?*

**S.P.:** Ich habe das Studium an der FU Berlin mit Informatik im Haupt- und BWL als Nebenfach begonnen. Nach dem Vordiplom bin ich – wie schon erwähnt – ein Jahr in die USA an die Duke University gegangen, um dort das Informatikstudium fortzusetzen. Nach meiner Rückkehr habe ich 2004 in Berlin meine Diplomprüfung abgelegt und mich dann auf Doktorandenstellen beworben. An der University of California, Berkeley, wurde ich – ebenfalls im Bereich Informatik – angenommen und habe dort promoviert. Der für die Promotion zunächst angedachte Bereich war Computer Vision, also visuelles Verstehen. Dies kam daher, dass ich mich im Zusammenhang mit den Fußballrobotern viel mit der Balllokalisierung beschäftigt hatte. Auch an der Duke University hatte ich mich besonders mit Gestenerkennung auseinandergesetzt, da wir dort Gesten der Blindensprache aufgenommen haben und versucht haben, sie zu entschlüsseln und mithilfe eines 3D-Modells auch zu generieren. Im Verlauf der Promotion habe ich dann ein immer größeres Interesse am Verstehen von Sprachen und Text im Kontext der künstlichen Intelligenz entwickelt und deshalb sogar den Doktorvater gewechselt. So kam ich zu meinem eigentlichen Promotionsthema, dem Textverstehen. In diesem Bereich bin ich bis heute tätig.

***Werner Weilhard:** Erinnern Sie sich noch an den Titel Ihrer Dissertation?*

**S.P.:** Ja, er lautet „Coarse-to-Fine Natural Language Processing“. Das war im Jahr 2009.

*W.W.: Eine Frage noch zur Sprache: Um in Amerika studieren zu können, muss man sehr gut Englisch können. Wo haben Sie das gelernt?*

**S.P.:** Am Arndt-Gymnasium, denke ich.

*W.W.: Aber nicht im Leistungskurs, sondern in Grundkursen?*

**S.P.:** Ja, ich denke schon, da ich keinen Englischunterricht außerhalb der Schule gehabt habe. Außerdem hat sich mein Englisch während des Studiums ständig verbessert, da in der Informatik viele Textbücher in Englisch verfasst sind. Zu weiteren Fortschritten kam es in den USA als Austauschstudent. Die Grundlagen wurden aber am Arndt-Gymnasium gelegt.

*B.S.: Das hören wir gern! In Ihrer Vita findet sich eine bemerkenswert große Anzahl von Veröffentlichungen. Welche davon erachten Sie selbst als besonders relevant?*

**S.P.:** Während meiner Promotionszeit habe ich in der Tat viel publiziert, was in der Forschungsabteilung bei Google, vor allem in den ersten Jahren, auch sehr geschätzt wurde. Meine allererste Publikation entstand während meines Austauschjahrs in Amerika und handelte von Gestenerkennung und Blindensprache. Das war mein erstes echtes Forschungsprojekt und hat mir die Tür für die spätere Promotion in den USA geöffnet. Meine erste wichtige Publikation in Berkeley war insofern spannend, als sie auf einem Vorlesungsprojekt basierte und mit einer Reise nach Australien zu einer Konferenz verbunden war, wo eine Präsentation stattfand. Deshalb war ich besonders motiviert. Mein Schwerpunkt lag damals noch im Bereich der Interpretation von Bildern. Bei der Publikation ging es dann aber vor allem um automatisierte Texterkennung: Um ein Beispiel zu nennen: Obwohl im Satzbau sehr ähnlich, besteht ein großer Unterschied zwischen den Sätzen „Ich esse ein Croissant mit Marmelade“ und „Ich esse ein Croissant mit der Gabel“. Mit Fragen, wie solche Unterschiede automatisiert erkannt werden können, habe ich mich sehr intensiv beschäftigt und viel Zeit investiert. Als Bewertung für die Publikation habe ich von drei verschiedenen Rezensenten „5/5“ erhalten, wusste jedoch zunächst nicht, was dies bedeutete, bzw. welche Skala dieser Bewertung zugrunde lag. Etwa eine von 5 bis 10 Punkten? Ich ging also zu dem Professor, und erhielt die Auskunft, dass die Skala nur bis fünf geht, die Bewertung also ganz ausgezeichnet sei, und er selbst ein solches Resultat bislang in vielen Jahren nur dreimal erhalten habe! Eigentlich lag mein Interessenschwerpunkt damals nach wie vor auf der Interpretation von Bildern, aber nach der Rückkehr von der Konferenz in Australien habe ich gespürt, dass das Thema der Erkennung von Sprache und grammatikalischen Strukturen von sehr hohem Interesse ist. Also habe ich von nun an in diesem Bereich weitergeforscht.

Bei Google ist mir eine Publikation, die am Anfang meiner Arbeit entstanden war, als sehr spannend in Erinnerung. Sie ist entstanden, als ich zum ersten Mal als „Advisor“ einen Praktikanten betreut habe. Wir haben sie dann gemeinsam zu einer Konferenz eingereicht und dafür den *Best Paper Award* erhalten. Dieser Praktikant ist dann übrigens zu Google gekommen und immer noch bei mir im Team. Nach diesem sehr schönen Start arbeiten wir nun schon 15 Jahre zusammen!

*B.S.: Werfen wir jetzt einen Blick auf Ihre berufliche Laufbahn. Welche wichtigen Stationen sind hier festzuhalten? Mir ist aufgefallen, dass Sie eine lange Zeit in den USA zugebracht haben. Aus welchen Gründen?*

**S.P.:** Wenn man sich meinen beruflichen Lebenslauf anschaut, ist da nur ein Eintrag: Ich bin fast seit 16 Jahren bei Google, davor war das Studium. Die Rolle bei Google hat sich aber in den Jahren sehr entwickelt. Nach meiner Promotion in Kalifornien habe ich bei Google in New York angefangen, nachdem ich dort schon ein Praktikum während der Promotion absolviert hatte.

*W.W.: Warum gerade bei Google?*

**S.P.:** Hier hatte ich die Möglichkeit, weiter zu forschen, aber gleichzeitig auch zu sehen, wie Forschungsergebnisse in Produkte umgesetzt – also praktisch genutzt – werden. Dieser Praxisbezug ist für mich sehr wichtig. Außerdem habe ich mir immer vorgestellt, dass ich irgendwann, nach ein paar Jahren in den USA, wieder zurück nach Europa möchte. Eine akademische Karriere in den USA kam für mich nicht in Frage, und zwar aus folgendem Grund: Strebt man dort an einer Uni eine Professur an, sind die ersten Jahre sehr schwer, und man steht unter immensem Druck, bis man es zur „*tenure*“, der vollen Professur, geschafft hat. Hat man sie dann, ist alles gut. Sie in den USA anzustreben, hätte aber für mich keinen Sinn gemacht, da ich ohnehin nicht vorhatte, dort zu bleiben. Ein Job in der Wirtschaft war deshalb für mich sinnvoller, zudem hat man – das habe ich schon im Praktikum so empfunden – bei Google sehr viele Freiheiten.

Viele der Produkte, die Google anbietet, basieren auf der automatisierten Interpretation von Texten. Es ist ein wesentliches Bestreben dieses Unternehmens, Informationen allgemein zugänglich zu machen. Die Forschung zur Interpretation von Texten ist hier besonders gefragt, zum Beispiel in Bezug auf die Beantwortung von Suchanfragen, Übersetzungen oder Situationen, in denen Benutzern beim Formulieren von Texten geholfen wird. So hat sich für mich bei Google eine sehr gute Mischung aus Arbeit an konkreten Produkten und Forschungstätigkeiten ergeben.

Zudem ist New York eine tolle Stadt und liegt außerdem näher an Europa als Kalifornien. Ein Wochenende in Europa – zum Beispiel um einen Geburtstag zu feiern oder eine Hochzeit zu besuchen, auf der man dann seine zukünftige Frau kennenlernen kann – ist

mit einer gewissen Anstrengung von New York aus möglich, von Kalifornien aus nicht mehr. Am Ende bin ich länger in New York geblieben als ursprünglich gedacht, nämlich zehn Jahre. Es hat Spaß gemacht, zumal ich an der New York University für sechs Semester neben der Tätigkeit für Google eine Vorlesung halten und dadurch auch meinen akademischen Weg weiterverfolgen konnte.

Meine Rolle bei Google hat sich stetig entwickelt: Erst habe ich viel programmiert, dann Praktikanten betreut, dann bin ich zum Manager geworden, habe ein kleines Team in New York geleitet, zu dem nach einer Restrukturierung ein weiteres in Kalifornien gekommen ist. Danach entschieden wir uns, zusätzlich noch ein Team in Seattle aufzubauen. Ich hatte also nach und nach an verschiedenen Orten verschiedene Teams, die ich von New York aus geleitet habe. Privat kam noch ein Glücksmoment hinzu, als ich – wie zuvor angedeutet – auf einer Hochzeit eines gemeinsamen Freundes in Frankreich meine zukünftige Frau kennengelernt habe. Sie ist dann nach New York gekommen, wo auch unser erster Sohn geboren wurde.

***B.S.:** Seit wann sind Sie wieder in Deutschland tätig? Welche Gründe gab es für diese Rückkehr?*

**S.P.:** New York macht ohne Kinder sehr viel Spaß und mit Kindern sehr viel Stress. Nach einigen Jahren dachten wir, es wird Zeit zurückzugehen. Deshalb war es eine glückliche Fügung, dass sich Google genau zu diesem Zeitpunkt zum Ziel gesetzt hatte, in Europa zu wachsen und der Meinung war, dass vor allem in Deutschland Handlungsbedarf bestehe. Einige haben sich damals dafür eingesetzt, nach München zu gehen. Ein Kollege und ich haben dem aber entgegengehalten, dass Berlin für die Anwerbung jüngerer Forscher attraktiver sein dürfte. Google schloss sich dieser Meinung an und bat uns, in Berlin beim Aufbau mithelfen. Das hat gut zu meinen Plänen gepasst, und so hat sich hier für mich eine spannende Tätigkeit ergeben. Ich hatte zwar auch einen Ruf auf eine Professorenstelle in Heidelberg, aber Heidelberg ist nun einmal nicht Berlin, und ich bin sehr mit dieser Stadt verbunden und fühle mich – wegen meiner Herkunft aus Bulgarien und des Wechsels nach Berlin im Kindesalter – sehr als Berliner. Würde ich in eine andere Stadt gehen wollen, kämen dafür vielleicht noch London oder Paris in Frage. Im Herbst 2019 sind wir also von New York nach Berlin umgezogen. Die Frage war dann, wie sich meine Rolle bei Google entwickeln würde, d.h. ob ich mich eher lokal fokussiere oder mein globales Team weiterleite. Mein Wunsch war, das Team zu behalten und von Europa aus zu steuern. Zum Glück bekam ich diese Möglichkeit. Sechs Monate später waren wir dann wegen der Corona-Pandemie ohnehin alle im Home-Office. D.h., ich hatte die richtige Entscheidung getroffen, denn lokal hätte ich hier gar nicht viel aufbauen können, weil praktisch alle potentiellen Arbeitskräfte zu Hause waren. Es hat also gar keinen Unterschied mehr gemacht, von wo aus ich arbeite. Hauptsächlich ist das Team in den USA und in anderen Ländern, und ich bin hier in Berlin. So hat sich ein

spezieller Arbeitsstil eingependelt, bei dem ich mit Videokonferenzen operiere, mitunter auch nach dem Abendessen. Morgens habe ich dafür etwas mehr Zeit, was bei kleinen Kindern sehr vorteilhaft ist.

***B.S.:** Würden Sie noch ein paar erläuternde Worte hinzufügen, um den Inhalt Ihrer Tätigkeit genauer zu beschreiben?*

**S.P.:** Sehr gerne. Seit mehr als 20 Jahren beschäftige ich mich mit dem automatisierten Verstehen und Interpretieren von Text. Ursprünglich ging es um die Erfassung grammatikalischer Strukturen mit dem Ziel, Übersetzungen zu verbessern. Wenn man zum Beispiel zwischen Deutsch und Englisch oder Deutsch und Französisch übersetzt, gibt es zwar keine riesigen Unterschiede in der Wortstellung, aber man muss dennoch zum Beispiel wissen, wo üblicherweise das Verb steht. Es ist also hilfreich, die grammatikalische Struktur zu kennen. Betrachtet man indes entferntere Sprachen, zum Beispiel Japanisch, können solche Fragen sogar eine entscheidende Rolle spielen, zum Beispiel, wenn das Verb immer hinten stehen muss. Anders als im Deutschen mit der üblichen Wortstellung „Subjekt, Prädikat, Objekt“ ist sie im Japanischen stets: „Subjekt, Objekt, Prädikat“. Es müssen hier also große sprachliche Distanzen überbrückt werden, um den Satzaufbau zu verstehen.

Derartige Überlegungen und Mechanismen dienten zuerst dazu, „Google Translate“ zu verbessern. Dann ging es, ungefähr in den Jahren 2013 und 2014, als „Siri“ und der „Google Assistant“ erschienen, um Befehle wie „Stelle mir den Wecker!“ oder Fragen wie „Wie ist das Wetter?“ Hierbei genügte es also nicht mehr, die Grammatik zu verstehen, sondern es ging auch um das Semantische, also zum Beispiel um die Frage, zu welcher Uhrzeit der Wecker gestellt werden soll. Es sollte möglich werden, solche Befehle oder Fragen auf verschiedene Art und Weise zu formulieren.

Im Team haben wir daran gearbeitet, entsprechende Systeme aufzubauen. Zunächst wurde die Google-Suche derart verändert, dass nicht alleine Links als Resultate ausgegeben werden, sondern dass Google auch Fragen beantwortet. Als wir damit anfangen, wurden diese Antworten existierenden Websites entnommen. Dies setzte zweierlei voraus: Einerseits musste der Sinn der Frage verstanden werden, und andererseits, worum es auf den entsprechenden Websites ging. Es musste also geklärt werden, ob die entsprechende Website zu der jeweiligen Frage passt.

Von uns entwickelte Systeme wurden dann auch für Werbeanzeigen auf Google verwendet. Dabei haben wir maschinelles Lernen bzw. künstliche Intelligenz genutzt, um die Systeme und Produkte zu verbessern. In den letzten 5-6 Jahren hat sich der Fokus von statistischen Modellen, bei denen zum Beispiel das Vorkommen des Wortes „book“ gezählt worden ist, hin zu Modellen verschoben, die in der Lage sind, verschiedene Bedeutungen (z. B. „etwas buchen“ als Verb oder „Buch“ als Nomen) zu unterscheiden. Hierfür kann zum Beispiel das Wort vor „book“ betrachtet werden. Ist es ein Artikel, wird es zu

fast 100 % ein Nomen sein. Stehen andere Wörter davor, können statistische Berechnungen weiterhelfen.

Aufgrund der Nutzung neuronaler Netze strebt man heutzutage Modelle an, die in der Lage sind, das nächste Wort passend vorherzusagen. Vereinfacht ausgedrückt, werden sie dadurch trainiert, dass man in einem Text, den man im Internet findet, den Anfang behält, das Ende entfernt und dann das Modell zur Vorhersage nutzt, welche Wörter anhand des bekannten Anfangs folgen werden. Je größer das Modell ist, desto besser wird es. In den letzten fünf Jahren haben wir daran gearbeitet, wie man solche Modelle ständig so vergrößert und trainiert, dass sie das nächste Wort zunehmend korrekt vorhersagen können. Dies ist der Bereich, mit dem ich mich gerade beschäftige. Als Ausgangsbasis dient ein Modell, das zwar bereits Wörter generieren kann, jedoch noch nicht so gut gesteuert werden kann. Der erste Teil ist das sog. „*pre training*“, der zweite Teil das „*post training*“ oder „*finetuning*“.

Mein Team ist dafür zuständig, dass man auf eine Frage hin eine korrekte Antwort erhält, dass zum Beispiel auf den Auftrag: „Schreib‘ ein Gedicht!“ hin wirklich ein Gedicht geschrieben wird. Verlangt der Auftrag auch noch einen bestimmten Rhythmus oder eine bestimmte Reimform, soll auch dies befolgt werden. Hierfür muss man entsprechende Beispiele für Inputs und Outputs sammeln und dann kreativ sein, weil man zwar nicht alle potentiellen Anfragen vorhersehen kann, aber möglichst viele davon abdecken möchte. Auch die verschiedenen Sprachen spielen hier eine Rolle. Das fragliche Modell könnte beispielsweise hauptsächlich auf Englisch trainiert werden, aber es bietet sich an, genügend multilinguale Beispiele hineinzugeben, damit es zum Beispiel auch auf Deutsch oder Französisch funktioniert, man aber trotzdem nicht die gesamte Arbeit, die man fürs Englische investiert hat, wiederholen muss. Da dies zu lange dauern würde, trainieren wir multilinguale Systeme, die Zugang zu allen Sprachen haben. Hierbei ist es bedeutsam zu evaluieren, ob sich eine sinnvolle Balance zwischen Aufwand und Ertrag ergibt. Teilweise geht es ja auch um Sprachen, die niemand im Team beherrscht. Es ist dann gar nicht leicht zu überprüfen, ob das System funktioniert. Wir sind hier an der Entwicklung und Verbesserung verschiedener Produkte beteiligt, um die Nutzbarkeit für Benutzer herzustellen.

***B.S.:** Im Mittelpunkt Ihrer Arbeit steht die KI. Wie kann man Grundprinzipien der künstlichen Intelligenz mit einfachen Worten Menschen ohne technische Vorkenntnisse bzw. ohne eine entsprechende Ausbildung erklären?*

**S.P.:** Ausgangsbasis sind zunächst die Daten, auf denen ein KI-Modell basiert. Man möchte „*pattern*“, also Muster in der Sprache, erkennen. Auf einem Bild kann man zum Beispiel einen Stuhl oder eine Blume erkennen. Muster können aber auch in Texten erkannt werden, und zwar dadurch, dass bestimmte Wörter häufig zusammen vorkommen. Das Modell versteht zwar nicht, was die zusammenhängenden Wörter bedeuten, kann

aber die Muster reproduzieren. Es ist wichtig, das Training so zu gestalten, dass nicht nur Aussagen generiert werden, die für das Modell plausibel sind, sondern, dass eine Kalibrierung erfolgt, damit das Modell weiß, wann es eine Antwort wirklich kennt und wann es hingegen besser ehrlicherweise angeben sollte, dass es eine Antwort nicht kennt.

Ein Beispiel hierfür ist, dass man das Modell nach einem Geburtsdatum einer beliebigen Person fragt, und es daraufhin eine (falsche) Antwort halluziniert. Richtig wäre hingegen zu sagen, dass die Antwort nicht bekannt ist. Wird hingegen nach dem Geburtsdatum einer besonders bekannten Person gefragt, darf das Modell die bekannte, richtige Antwort angeben. Das Erstellen eines Entwurfs für einen Satzesatz oder das Ausformulieren eines Satzesatzes anhand von vorgegebenen Stichwörtern funktionieren bereits gut. Man muss jedoch aufpassen, dass das Modell nicht falsche oder unpassende Informationen hinzufügt, die in den ursprünglichen Stichwörtern nicht enthalten sind oder keinen Sinn ergeben, für das Modell jedoch plausibel sind.

Hieran wird sehr viel geforscht, da es nicht leicht ist, hierüber die vollständige Kontrolle zu haben. Bevor „ChatGPT“ erschien, hat Google deshalb auch keine Produkte mit solchen Modellen gebaut. Wir waren damals noch nicht sicher, auf der richtigen Seite zu stehen, solange wir keine komplette Kontrolle über die Modelle haben und deshalb die Abwägung von Nutzen und Risiko nicht eindeutig ausfällt. Als „ChatGPT“ dann auf dem Markt war, erhöhte sich natürlich der Druck, auch ein ähnliches System anzubieten. Die Erwartungshaltung der Benutzer hatte sich verändert. Wäre Google zuerst mit einem solchen Modell an die Öffentlichkeit getreten, wäre beim Auftreten von Fehlern die Akzeptanz der Nutzer möglicherweise nicht so hoch gewesen. Bei „ChatGPT“ war die Toleranz deshalb höher, weil man einem Start-up eine höhere Fehleranfälligkeit zugestanden und auch verziehen hat. So wurde dann wahrgenommen, dass ein solches Modell zwar ab und zu Fehler macht, aber vor allem, dass es mindestens unterhaltsam und vielleicht sogar nützlich ist. In den ungefähr zwei Jahren, in denen es „ChatGPT“ mittlerweile gibt, sind wir auf der Nützlichkeitskala in jedem Fall vorangekommen. Das Modell „halluziniert“ zwar immer noch, und man muss vorsichtig sein, aber es kann mittlerweile die Arbeit erleichtern und in den verschiedensten Bereichen die Produktivität steigern.

***B.S.:** Halluzinieren bedeutet in diesem Zusammenhang: etwas hinzuzudichten, das zwar für das Modell anhand der Wortfolge plausibel, aber nicht unbedingt von der Information her korrekt ist?*

**S.P.:** Genau. Halluzinieren kann zum Beispiel passieren, wenn das Modell etwas nicht wissen kann: Wenn man fragt, wie Hertha BSC in der letzten Woche gespielt hat, das Modell aber mit Daten aus vorherigen Monaten trainiert wurde, kann es sein, dass es etwas halluziniert. Die Modelle wissen eben nicht, was sie nicht wissen können. Wir versuchen durch das Training, solche Situationen zu vermeiden, d.h. ,entweder muss das Modell die Aktualität der Frage erkennen und sagen, dass es die Antwort nicht kennt, oder

eben eine Suchanfrage stellen und die Ergebnisse zusammenfassen oder daraus die richtige Antwort extrahieren. Hierfür muss also ein Zugriff auf eine Suchmaschine gegeben sein. Wir arbeiten daran, dass die Modelle erkennen, wann solche zusätzlichen Möglichkeiten – wie der Zugriff auf eine Suchmaschine – erforderlich und möglich sind und wann nicht. In letzterem Falle muss dann die Antwort eben lauten: „Ich weiß es nicht.“

**B.S.:** *Kommen wir nun zu einigen Fragen, die mit KI zusammenhängen: Wie viele wesentliche oder große KI-Modelle gibt es momentan?*

**S.P.:** Es sind viele, und es werden immer mehr. Von OpenAI gibt es ja außer „ChatGPT“ noch weitere Modelle. Dann gibt es Anthropic aus den USA, die haben auch gute Modelle wie „Claude AI“ entwickelt. Ebenso Google mit „Gemini“. Elon Musks Unternehmen X.AI hat kürzlich das neue Modell „Grok“ vorgestellt, das bei Twitter/X eingebaut ist. Dies sind alles proprietäre, also geschlossene Modelle, auf die man zwar zugreifen kann, bei denen aber der Quellcode nicht bekannt ist und von denen man auch nur wenige technische Details kennt. Daneben gibt es sog. Open Source-Modelle, also offene, bei denen der Code und die Gewichtungen innerhalb der Modelle öffentlich sind. Hierzu zählen beispielsweise die Modelle „Llama“ von Meta, „Gemma“ von Google und die des französischen Start-up-Unternehmens Mistral AI. Alle diese Modelle sind etwas kleiner. Man kann sie sich herunterladen und sogar selbst modifizieren. Die kleineren Varianten können auf einem Smartphone laufen, für die größeren benötigt man einen Computer. Es ist wichtig, diese beiden Klassen an Modellen zu unterscheiden. Aktuell stellt sich die Frage, ob sich die großen, proprietären Modelle durchsetzen werden oder die kleineren, veränderbaren, offenen Systeme. In beiden Fällen bestehen sie aus Milliarden von Parametern, sogenannten Gewichten, wobei die einzelnen Gewichte für sich selbst genommen jeweils keine klar definierbare Bedeutung haben. Man kann also die kausalen Zusammenhänge innerhalb solcher Modelle nicht nachvollziehen. Wichtig ist, mit welchen Daten sie trainiert worden sind. Auch bei den offenen Modellen ist dies nicht immer ganz klar, obwohl es hierzu teilweise Listen mit Millionen von Wörtern gibt.

**B.S.:** *Wie ordnen Sie in diesem Zusammenhang „DeepSeek“ aus China ein?*

**S.P.:** Dieses Modell hat ja seit einiger Zeit sehr viel Aufmerksamkeit bekommen, weil es zwar klein, aber doch sehr performant ist. Es wirft daher die grundsätzliche Frage auf, ob die großen Modelle wirklich benötigt werden. Sie verbessern zwar die Qualität, aber der Zusammenhang zwischen Größe und Qualität ist logarithmisch, nicht linear. Das heißt: Irgendwann benötigen selbst kleine Verbesserungen der Qualität so erhebliche Vergrößerungen der Modelle, dass man hinsichtlich der Menge der Trainingsdaten und der Rechenleistung an Grenzen stößt.

Seit einiger Zeit sehen wir ein starkes Aufholen der kleineren Modelle. Was also im letzten Jahr gemäß den Standards die Qualität des besten großen Modells war, kann in diesem Jahr schon mit einem Modell erreicht werden, das 10-50-mal kleiner ist. So hätte es noch vor zwei oder drei Jahren keinen Sinn gemacht, KI-Modelle in ein Smartphone zu integrieren. Da nun aber die kleineren Modelle deutlich besser geworden sind und die Smartphones mehr Rechenleistung haben, können jetzt Modelle wie „ChatGPT“, die noch Kurzem nur in der Cloud lauffähig waren, auch auf einem Smartphone betrieben werden. Hierdurch werden neue Möglichkeiten eröffnet, zum Beispiel in Bezug auf die Privatsphäre; zudem wird eine direkte Einbindung in Apps ermöglicht.

***B.S.:** Welche Kriterien sind beim Vergleich von KI-Modellen relevant, und wodurch zeichnet sich ein besonders gutes KI-Modell aus?*

**S.P.:** Das sind Fragen, mit denen wir uns intensiv beschäftigen, da sie für die Forschung wirklich relevant sind. Bereits heute können KI-Modelle in vielen Bereichen mehr leisten als Menschen. So meistern sie zum Beispiel Tests in Physik oder in der Medizin besser als Studierende am Ende ihres Studiums. Ähnliches erscheint auch in vielen anderen Fächern möglich. Bei vielen generierten Antworten müssen deshalb schon jetzt selbst Experten ganz genau hinschauen, um zu bemerken, dass sie künstlich erzeugt worden sind. Es wird deshalb immer komplizierter, die Modelle zu evaluieren, weil es schwer ist, valide Testkriterien zu definieren und sich ständig neue Tests auszudenken. Es ist wie in der Schule, wo es auch schwer ist, sinnvolle Klausuren zu entwickeln, bei denen die Resultate eine hohe Aussagekraft haben. Um neue Tests zu entwickeln und valide Bewertungskriterien festzulegen, arbeiten wir deshalb mit Universitäten zusammen. Erfreulicherweise gelangen wir dabei zu schnellen Fortschritten. Weshalb ist das so? Auch wenn man Modelle nicht speziell auf die Tests hin trainiert, fließen irgendwann die Informationen aus den Tests indirekt in die Modelle selbst ein. Folglich werden sie sukzessive sehr gut in den Tests abschneiden, weil sie indirekt dafür optimiert worden sind – zum Beispiel dadurch, dass mehrfach derselbe Test verwendet worden ist und Resultate der Tests für Optimierungen des Modells genutzt worden sind. Dann müssen neue Aufgaben in den Tests her, da die alten nicht mehr repräsentativ sind.

Früher hat es Jahre gedauert, bis ein Modell derart optimiert war, dass es die Tests gut lösen konnte. Heute ist dies nur noch eine Frage von Monaten. Ist das Modell besser als Menschen, muss man sich neue, schwierigere Aufgaben überlegen. Solche Vorgänge sind sehr zeitaufwändig und teuer, aber auch wichtig. Wir arbeiten intern sehr intensiv an der Kreation neuer Evaluationen und veröffentlichen diese dann auch, damit objektive Vergleichskriterien für verschiedene Modelle geschaffen werden. Kopien dieser Aufgaben verbreiten sich schnell, so dass Aufgaben und deren Lösungen wiederum Eingang in die Modelle finden, da diese ja mit öffentlich zugänglichen Daten trainiert werden. Zwar freut man sich dann durchaus über gute Resultate, denn scheinbar hat man ja eine

Verbesserung erreicht. In Wahrheit jedoch hat das Modell sozusagen „geschummelt“: Wenn nämlich die Aufgaben im Internet stehen, hat es sie schon während des Trainings kennengelernt und konnte sie deshalb sehr gut lösen. Aus diesem Grund müssen jedes Jahr neue Tests entwickelt werden.

*W.W.: Wie sieht denn das Training konkret aus, was wird da im Detail gemacht?*

**S.P.:** Den größten Anteil macht das „pre training“ aus, bei dem einfach Daten aus dem Internet, die öffentlich zugänglich gemacht wurden, gesammelt werden. Sie werden zunächst in einer großen Textdatei aufbereitet. Dann wird das Modell trainiert, indem man das erste Wort vorgibt und dann das Modell das zweite Wort vorhersagen soll. Sodann gibt man die ersten beiden Wörter vor und lässt das dritte Wort vorhersagen etc. In dieser Art und Weise schreitet man weiter voran: Man gibt also eine Anzahl an Wörtern vor und schneidet nachfolgende Wörter ab. Auf diese Weise wird das Modell trainiert, die Sätze zu vervollständigen.

*W.W.: Und das Trainingsinstrument sind Algorithmen?*

**S.P.:** Genau, man passt nach jedem Wort die Parameter des Modells so an, dass das nachfolgende Wort, das wirklich gesucht war, die höchste Wahrscheinlichkeit erhält, vom Modell vorgeschlagen zu werden. Ein solches Vorgehen wiederholt man iterativ, bis alle Trainingsdaten durchlaufen worden sind.

*B.S.: Das sind die Gewichtungen, von denen Sie schon gesprochen haben?*

**S.P.:** Ja, genau, hier wird es mathematisch. Man kann Formeln aufstellen, die diese Berechnungen übernehmen.

*B.S.: Wie wichtig ist der stetige Anstieg von Rechenleistung für die künstliche Intelligenz?*

**S.P.:** Die Entwicklung der Rechenleistung ist einer der wichtigsten Faktoren, die den Fortschritt vorantreiben. Viele der Modelle oder Grundideen gab es schon früher, man konnte sie jedoch nicht umsetzen, da ein Training aufgrund des Mangels an Rechenleistung nicht möglich war. Also behalf man sich mit Approximationen und Vereinfachungen oder mit weniger genauen oder einfach kleineren Modellen. Zum Beispiel war während der Zeit meiner Promotion ein Modell mit einer Million Parametern schon riesig. Jetzt haben wir Modelle mit hunderten von Milliarden von Parametern. Das ist möglich geworden durch die heute zur Verfügung stehende Rechenleistung und auch durch gezielte Investitionen.

*B.S.: Bedeutet das: Grundlage eines jeden KI-Modells sind riesengroße neuronale Netze?*

**S.P.:** Ja, und für das Training dieser Netze werden vorrangig die hierfür besonders geeigneten Grafikkarten bzw. Chips der Nvidia Corporation genutzt. Dieser Umstand hat einen Großteil des Fortschritts ermöglicht. Von den Algorithmen her hat sich natürlich auch vieles getan: Wir haben derweil ein tieferes Verständnis, wie man die Parameter (also die Gewichte) bestmöglich anpasst.

*B.S.: Hierdurch konnte also die Qualität von neuronalen Netzen weiter optimiert werden?*

**S.P.:** Richtig, und hierin besteht der hauptsächliche Unterschied zwischen den unterschiedlichen Modellen der verschiedenen Firmen. Es gibt zwar auch geringfügige Unterschiede in der Architektur, aber die größten liegen darin, welche Daten für Trainingszwecke genutzt worden sind, wie genau Optimierungen vorstättengingen und wie gut das Entwicklerteam darin war, sie zu finden.

*B.S.: Welchen Einfluss werden Ihrer Meinung nach Quantencomputer auf die Entwicklung künstlicher Intelligenz haben?*

**S.P.:** Hierzu muss ich sagen, dass ich es sehr schwer finde, mich in Quantencomputer und deren Funktionsweise hineinzudenken. Ich weiß aber, dass es möglich ist, damit manche Probleme, deren Lösungen mit traditionellen Computern extrem lange dauern würden, schnell zu bewältigen. Um dies am berühmten Beispiel des „traveling salesman“ zu demonstrieren: Dieser möchte eine bestimmte Anzahl an Städten hintereinander besuchen und hierfür den optimalen Weg – also den kürzesten oder den schnellsten – bestimmen. Jedes Mal, wenn man eine neue Stadt hinzufügt, wird das Problem exponentiell schwieriger, da sich sehr schnell viele weitere Kombinationsmöglichkeiten ergeben. Ein traditioneller Computer geriete hier irgendwann, wenn man Tausende Städte berücksichtigen wollte, an seine Grenzen. Quantencomputer hingegen können gleichzeitig eine Vielzahl von Kombinationen ausprobieren, da sie – im Gegensatz zu normalen Computern – mehr als zwei Zustände (0 und 1 bzw. „An“ und „Aus“) kennen. Bei Quantencomputern existieren Quantenstadien, die sich *zwischen* den Zuständen „An“ und „Aus“ befinden. So werden Probleme wie das oben genannte, die eigentlich mit zunehmender Komplexität exponentiell rechenaufwändiger werden, lediglich linear rechenaufwändiger. Dies hat immense Implikationen, zum Beispiel für die Kryptologie. Passwörter könnten zum Beispiel viel einfacher geknackt werden. Implikationen für die künstliche Intelligenz könnten aber auch dadurch gegeben sein, dass eine optimale Lösung sofort gefunden werden kann, ohne iterative Prozesse durchführen zu müssen. Ich glaube allerdings, dass

wir hiervon noch einige Jahre entfernt sind, wenngleich es auf jeden Fall deutliche Fortschritte in diesem Gebiet gibt. Ganz besonders spannend ist in diesem Kontext Folgendes: Wir nähern uns dem Zeitpunkt, ab dem wir in der Lage sein werden, die KI selbst zur Weiterentwicklung der KI zu nutzen, d.h., wir nutzen sie selbst zum Auffinden der optimalen Gewichtungen in den neuronalen Netzen. Hierdurch könnte auch die Integration von Quantencomputern beschleunigt werden.

*B.S.: Das heißt, wir stehen nach wie vor am Anfang einer exponentiellen Entwicklung, was die Qualität, die Geschwindigkeit und auch die Größe der neuronalen Netze und auch der künstlichen Intelligenz angeht?*

**S.P.:** Ja, das sagen auf jeden Fall die Optimisten, und ich zähle mich eher dazu. Aber es gibt auch, denke ich, eine nicht triviale Wahrscheinlichkeit von vielleicht 5-10 %, dass irgendetwas Fundamentales noch fehlt, dass also die Architekturen, die wir haben und die Art und Weise, wie wir die Netze trainieren, noch zu vereinfacht sind. Prof. Jitendra Malik, anfangs mein Doktorvater in Berkeley, hat öfter gesagt: „Nur weil man einen Berg hinaufsteigen kann, heißt das nicht, dass man mit der gleichen Technik bis zum Mond kommt.“ Die Frage ist also: Kommen wir mit den gleichen Techniken und immer mehr Rechenleistung wirklich immer weiter, oder nähern wir uns einem Plateau, so dass eine fundamentale Änderung zum Erreichen der nächsten Schritte notwendig ist? Das weiß momentan niemand.

*B.S.: Ich möchte gerne noch einmal einen Aspekt des chinesischen KI-Modells „DeepSeek“ ansprechen. Aufsehen hat es ja u.a. dadurch erregt, dass es angeblich mit besonders geringem Aufwand trainiert wurde. Wie beurteilen Sie dies?*

**S.P.:** „DeepSeek“ ist ein sehr gutes kleines und kostengünstiges Modell, aber die Informationen darüber sind insofern etwas irreführend, als die Kosten für nur einen einzigen Trainingslauf angegeben worden sind. In der Praxis ist es aber nicht so, dass man auf Anhieb eine gute Lösung findet, sondern man probiert mehrere Trainingsläufe aus und nähert sich einem guten Ergebnis iterativ. Die Kosten sind also wahrscheinlich deutlich größer als angegeben, aber immer noch kleiner als bei großen Modellen.

Der zweite Faktor, den wir nicht genau kennen, ist die Frage, auf welcher Datenbasis „DeepSeek“ trainiert wurde. Beim „pre training“ machen die meisten Entwickler dasselbe und nutzen öffentliche Daten. Die großen westlichen Firmen müssen sich dabei jedoch an Gesetze halten. Wenn also beispielsweise die New York Times sagt, dass mit ihren Daten nicht trainiert werden darf, dann machen wir das auch nicht. Die diesbezügliche Vorgehensweise von „DeepSeek“ kennen wir nicht.

Weiterhin ist es ein sehr aufwändiger Prozess, menschlich erstellte (Text-)Beispiele zu sammeln und als Trainingsdaten aufzubereiten bzw. zur Verfügung zu stellen. Man kann

dafür aber auch vorhandene KI-Modelle einfach kopieren und daraus Trainingsdaten herstellen, die man häufig nutzt. Dies wäre aber Schummeln. Ich will damit nicht unterstellen, dass die Entwickler von „DeepSeek“ so vorgegangen sind, dennoch stellt sich die Frage, ob von anderen Modellen kopiert worden ist, wie manche Branchenkenner vermuten. Das Modell „DeepSeek“ als solches funktioniert aber sehr gut.

***B.S.:** Vor kurzem wurde „DeepSeek“ in Südkorea verboten. Halten Sie derartige Verbote für sinnvoll?*

**S.P.:** Grundsätzlich ist es sehr schwierig, Verbote im Internet umzusetzen, da sich meistens schnell ein Umgehungsweg findet. Ich mache mir allerdings schon Sorgen, in welchem Zustand manche KI-Modelle veröffentlicht werden, weil sie eben auch zu Zwecken der Desinformation genutzt werden können. Denn: Auch ein Modell, das in vielen Benchmarks und gemäß vieler Regularien gut abschneidet, kann ja trotzdem in subtiler Art und Weise manipuliert worden sein, so dass es beispielsweise Propaganda für bestimmte politische Richtungen macht und im Sinne des Betreibers antwortet. Hierfür fehlt uns, glaube ich, noch das rechte Verständnis, auch in der Politik. Es ist gewiss ein Problem, dass noch keine allgemein anerkannte Regulierung besteht, ja, dass wir noch nicht einmal wissen, wie derartige Regulierungen und hierfür notwendige Evaluationen aussehen könnten. Die Entwicklung der KI schreitet sehr schnell voran, und im Augenblick ist noch ziemlich unklar, was wir auf diesem Gebiet überhaupt brauchen.

Bei Google gibt es in Bezug auf die Suche ein dickes Buch, knapp 200 Seiten stark, mit Kriterien, wie Suchresultate bewertet werden. Dieses Buch ist öffentlich, d.h., alle können es einsehen. Wir haben Mitarbeiter, die dieses Buch praktisch auswendig kennen und beurteilen können, ob Änderungen an der Suchmaschine vorteilhaft sind oder nicht. Es war sehr viel Arbeit vonnöten, diese Kriterien so objektiv wie möglich zu gestalten. Die Suchresultate sollen relevant sein, aber auch Diversität zum Ausdruck bringen, also verschiedene Perspektiven wiedergeben, so dass keine „Filterbubble“ entsteht, in der in allen Suchresultaten die gleiche Meinung vertreten wird. Hier gibt es allerdings bestimmte Grenzen, zum Beispiel beim Holocaust, bei dem Google selbstverständlich eine wissenschaftliche und historisch korrekte Perspektive einnimmt.

Grundsätzlich versucht Google also, sich an wissenschaftlichen Resultaten zu orientieren. Bei KI-Modellen gibt es jedoch bislang noch keine solchen allgemein anerkannten Prinzipien oder Richtlinien. Jeder Hersteller eines KI-Modells versucht, hier so gut wie möglich eine vernünftige Lösung zu finden. Aber da ein großer Zeitdruck besteht, werden zum Teil Modelle veröffentlicht, die noch nicht ganz ausgereift sind. In Bezug auf mögliche gesellschaftliche Folgen sollte man hier in jedem Fall aufpassen. Verbote halte ich hingegen in diesem Bereich eher für wenig sinnvoll, da man sie, wie gesagt, immer irgendwie umgehen kann.

*W.W.: Was ist in dem Zusammenhang zum Beispiel von dem „EU AI Act“ zu halten?*

**S.P.:** Im Großen und Ganzen begrüße ich diese Initiative. In der Praxis ist indes noch kein großer Einfluss auf KI-Modelle zu verspüren. Betroffen sind eher Anwendungen, bei denen im Fall von Fehlfunktionen direkt Menschenleben gefährdet sind. Die KI-Modelle, von denen wir hier sprechen, könnten jedoch bei Fehlfunktionen ebenfalls zumindest indirekt Menschenleben gefährden, insbesondere hinsichtlich möglicher langfristiger Folgen. Trainingsdaten müssen deshalb veröffentlicht werden, aber ich weiß nicht, inwieweit es wirklich hilfreich sein wird, zum Beispiel eine Liste mit Billionen oder Trillionen von Wörtern zu veröffentlichen, die zum Training genutzt worden sind. Eine solche Liste kann niemand durchlesen und niemand verstehen. Deswegen befinden wir uns hier aus meiner Sicht im Bereich der Bürokratie, ohne dass der gewünschte Zweck wirklich erreicht wird. Ich würde mich über eine effiziente Regulierung, zu der man Vertrauen haben kann, sehr freuen. Für KI-Modelle wäre eine Regulierung ähnlich wie beim TÜV wünschenswert, bei dem man genau erfährt, welche Mängel am Auto behoben werden müssen. Sind diese dann behoben, kann man sich auch sicher sein – zum Beispiel, dass das Auto bremsen wird, wenn gebremst werden muss. Für KI-Modelle existiert eine solche Regulierung bisher leider noch nicht.

*B.S.: Wie beurteilen Sie in diesem Zusammenhang das Stargate Projekt der Trump-Administration?*

**S.P.:** Mir scheint es, dass in der Politik häufig große Ankündigungen gemacht werden, wobei jedoch fraglich ist, wie viel Substanz wirklich dahintersteckt. Ich kann mir gut vorstellen, dass in den USA massiv investiert werden wird. Es gibt jedoch große Zweifel, ob wirklich 500 Milliarden \$ zur Verfügung stehen werden. Nichtsdestotrotz sollte das Projekt ein Weckruf für Europa sein, mehr in die eigene digitale Infrastruktur zu investieren. Die dafür notwendigen Summen sind sehr groß, aber die EU insgesamt ist auch ein großer Wirtschaftsraum und sollte deshalb mithalten können. Ich fürchte allerdings, es wird aktuell eher gekleckert als geklotzt!

*B.S.: Das führt unmittelbar zur nächsten Frage: Häufig wird behauptet, dass Europa - und damit auch Deutschland - bei der Entwicklung von KI gegenüber China und den USA abgehängt würden. Wie sehen Sie diese Problematik, und was müsste in Europa und speziell in Deutschland geschehen, um wieder Anschluss zu finden?*

**S.P.:** Ich sehe uns nicht als abgehängt, weder was die Forschung noch was das Personal betrifft. Es gibt sehr gute deutsche und europäische Forscher\*innen – viele davon in den USA –, mit denen ich jetzt im Rahmen von Google DeepMind zusammenarbeite. Google DeepMind ist 2023 durch die Zusammenlegung mit Google Brain entstanden, nachdem

Google das britische Start-up „DeepMind Technologies“ mit Sitz in London bereits 2014 übernommen hatte. Ich kann mir gut vorstellen, dass die Ansiedlung neuer Firmen in Deutschland dazu führen könnte, dass viele Forscher\*innen hierher (zurück)kommen. In Berlin gibt es bereits einige Start-ups im Bereich der künstlichen Intelligenz, die ich indes momentan als technologisch noch nicht ganz auf dem letzten Stand einordnen würde. Einer der Gründe hierfür sind bürokratische Hemmnisse, die zu Zeit- und Geldverlusten führen. Dies erzeugt nicht gerade ein investitionsfreundliches Klima. Um einen Mehrwert zu generieren, muss aber gerade im Bereich von KI viel Wagniskapital aus der privaten Wirtschaft beschafft werden, denn der Staat selbst ist ja keiner der größten Investoren – übrigens auch bei Stargate nicht! Um zu erkennen, was konkret getan werden kann, damit talentierte junge Menschen im Land bleiben und sie davon abzuhalten, in andere Länder gehen, um Firmen zu gründen, hilft ein Blick nach Frankreich. Präsident Macron hat seit Beginn seiner ersten Präsidentschaft viel über KI gesprochen und Gesetze veranlasst, die Firmengründungen in diesem Bereich fördern. Im Ergebnis gibt es nun in Paris einige Start-ups, die wirklich auf Weltniveau sind. Ich wünsche mir das auch für Deutschland.

***B.S.:** In welchen Lebensbereichen werden Ihrer Meinung nach Auswirkungen der künstlichen Intelligenz besonders spürbar sein? Pars pro toto seien folgende Bereiche genannt: Gesundheit und Medizin, Arbeitsmarkt, Kunst und Kultur, Informationen – zum Beispiel auch das Thema „Fake News“ – und das Militär. Welche weiteren Bereiche könnte man in Betracht ziehen?*

**S.P.:** Ich glaube tatsächlich, die KI wird in allen Bereichen relevant sein. Selbst Menschen, die der Meinung sind, dass die KI sie nicht berühren wird, werden irgendwann davon betroffen sein, da sich vieles verändern wird. Voraussagen in diesem Bereich sind schwierig und haben sich in der Vergangenheit häufig als falsch erwiesen. So dachte man lange, dass Schach und andere intellektuelle Spiele die schwierigsten Herausforderungen für eine künstliche Intelligenz wären. Dann hat vor ungefähr 20 Jahren der Computer *Deep Blue* Weltmeister Garri Kasparow besiegt. Wir lernten also, Computer können Schach spielen! Inzwischen ist diese Entwicklung weitergegangen und erstreckt sich auch auf andere Brettspiele wie Go.

Man war lange auch der Meinung, dass die Erstellung von Kunst für die künstliche Intelligenz schwierig sei. Ich glaube, dass es hier natürlich nicht nur um das geht, was man sieht oder hört, sondern um den künstlerischen Gedanken, der einem Kunstwerk zugrunde liegt. Einen solchen künstlerischen Gedanken gibt es bei der KI nicht. Trotzdem wird man künstlich generierte Werke in der Malerei oder der Musik bald nicht mehr von Kunstwerken unterscheiden können, die von Menschen erschaffen wurden. Daher auch meine Annahme, dass die KI alle Lebensbereiche betreffen wird.

Die Herausforderung wird darin bestehen zu erkennen, wie die KI kreativ – also im Zusammenspiel mit Menschen – genutzt werden kann. Zum Beispiel können viele

Menschen – mich eingeschlossen – nicht gut zeichnen. Eine interessante Idee würde also vielleicht nie zeichnerisch umgesetzt werden, sondern für immer ein Gedanke bleiben. Mit der KI wird man hingegen die Möglichkeit haben, mit Wörtern zu beschreiben, was man visuell darstellen möchte. Hierauf kann eine iterative Zusammenarbeit mit dem Computer folgen, durch die eine visuelle Darstellung des gedanklichen Konzepts real umgesetzt werden kann. Der Computer kann hier also Werkzeug dienen, d.h., die spezifischen Fähigkeiten von Künstler\*innen könnten in der Zukunft weniger exklusiv sein. Eine solche Entwicklung sehe ich auch in der Medizin. Am Anfang wurde häufig gesagt: „Die Maschine kann es besser als ein Arzt.“ Aktuell heißt es eher: „Die Maschine ist ein nützliches Werkzeug für den Arzt“. Diese Entwicklung halte ich für positiv. Letztendlich muss ein Arzt ein Gespräch mit dem Patienten führen und eine gute Entscheidung treffen. In der Praxis ist es jedoch schwer, 40 oder mehr Stunden zu arbeiten und dann noch die neuesten wissenschaftlichen Erkenntnisse zu studieren. Mit den Werkzeugen der KI können sich Ärzte jetzt auf unkomplizierte Weise auf den neuesten Stand bringen, was ihnen hoffentlich auch gestatten wird, mehr Zeit mit Patienten zu verbringen. Es ist allerdings auch davon auszugehen, dass manche Berufe durch KI obsolet werden. Dies dürfte in allen Disziplinen zu beobachten sein.

*W.W.: Im Bereich des Militärs wäre ja insbesondere eine Abschätzung der Risiken von zentraler Bedeutung. Würde man die KI sozusagen omnipotent oder universell einsetzen, also ohne Restriktionen, könnte das fatale Folgen haben. Welche Risiken kann man hier benennen, und wie könnte man sie möglicherweise einschränken?*

**S.P.:** Ich muss gestehen, dass ich mich mit dem Aspekt des Militärs nicht viel beschäftigt habe, insbesondere weil ich in Bezug auf die künstliche Intelligenz nicht an militärischen Sachverhalten arbeiten möchte. Mir ist aber durchaus bewusst, dass dies jemand tun wird bzw. dass dies bereits geschieht. Ich denke, hier müssen unbedingt sowohl das Kriegs- als auch das internationale Völkerrecht erweitert werden, um entsprechende Prinzipien und Richtlinien zu schaffen. Sich auf Regulierungen zu einigen, halte ich aber in Anbetracht der aktuellen Weltlage leider für sehr schwierig. Hier mache ich mir schon Sorgen, denn KI-Modelle können ja auch systematische Fehler machen und zum Beispiel bestimmte Gruppen oder Minderheiten benachteiligen. Deshalb halte ich ihre Nutzung ohne Regulierung besonders im militärischen Bereich für sehr gefährlich.

*W.W.: Zumal hier ja auch der Benutzer des KI-Modells eine entscheidende Rolle spielt. Handelt es sich um demokratisch gewählte und legitimierte Regierungen, ist die Sorge vielleicht etwas geringer. Aber wenn wir z. B. an Russland oder China denken, muss man sich schon mehr Sorgen machen.*

**S.P.:** Als Wissenschaftler überlege ich mir sehr wohl, für wen ich arbeite. Es ist mein Bestreben, dass „gute“ Regierungen über die Möglichkeiten der KI verfügen können. Dass die meisten Modelle jedoch heutzutage von amerikanischen Firmen entwickelt werden, bereitet mir ebenfalls Sorgen – mehr noch als vor ein paar Jahren, weil da noch ein enges wertebasiertes Bündnis zwischen den USA und Europa existierte. Heutzutage ist dies so nicht mehr gegeben, auch deshalb wünsche ich mir Alternativen für Europa.

**B.S.:** *Wird die künstliche Intelligenz sich selbst automatisiert weiterentwickeln können, oder geschieht dies bereits? Welche Erwartungen haben Sie in dieser Hinsicht?*

**S.P.:** KI-Systeme haben kein Bewusstsein und werden sich nicht selbst Ziele setzen können. Ich lese allerdings gerade ein spannendes Buch von Yuval Noah Harari mit dem Titel „Nexus“, das letztes Jahr erschienen ist und schöne Beispiele dafür aufzeigt, wie KI-Systeme jetzt schon autonom funktionieren und Algorithmen entscheiden, was bei Instagram oder Facebook gezeigt wird. Was bedeutet das? Der Mensch hat zwar das Ziel gesetzt, und die Modelle können die Erreichung des Ziels optimieren, aber wenn der Mensch nicht alle Implikationen durchdacht hat, können die Systeme irgendwelche Abkürzungen finden, die zum Teil nicht gewollt sind.

Ich bin der Überzeugung, dass es solche Vorgänge geben wird. Die KI wird immer größere Aufgaben übernehmen, und der Mensch wird sich auf autonome Systeme verlassen, ohne zu verstehen, was diese Systeme jeweils genau machen. Die Auswirkungen werden den Menschen dann erst später bewusst werden. Heute kann der Mensch zwar immer noch den Stecker ziehen, die künstliche Intelligenz hat sich also noch nicht verselbstständigt, aber man hat ihr bereits sehr viele Möglichkeiten gegeben, und sie wird vielleicht Lösungen finden, die man als Mensch nicht befürwortet, jedoch ohne der KI gesagt zu haben, dass man diese Lösungen nicht gut findet. Vielleicht ist es dann schon zu spät, denn ist der Schaden erst einmal angerichtet, ist es schwierig, ihn zu beheben.

**B.S.:** *Zum Abschluss dieses Gesprächs möchten wir Sie nach Ihren persönlichen Wünschen für die zukünftige Nutzung von KI fragen.*

**S.P.:** Ich hoffe, dass wir uns auf das Positive fokussieren und KI benutzen, um medizinische Fortschritte voranzutreiben, um Krankheiten zu heilen und um den Wohlstand zu verbessern – und nicht, um Gesellschaften weiter zu spalten, sei es in ökonomischer oder in politischer Hinsicht. Hierfür muss noch viel gearbeitet werden. In Bezug auf die Forschung müssen wir beispielsweise noch besser verstehen, was wirklich innerhalb eines KI-Modells geschieht und wie wir solche Modelle besser kontrollieren können. Vorteilhaft wäre es auch, wenn die Nutzer ein tieferes Verständnis von KI-Modellen erlernen würden, um deren Nebenwirkungen und Risiken besser einschätzen zu können.

Kritisches Hinterfragen ist dafür eine wesentliche Voraussetzung. Ich habe aber das Gefühl, dass wir in einem Zeitraum vom Mauerfall bis etwa zum Ende des 1. Jahrzehnts in diesem Jahrhundert in einer sehr optimistischen Phase waren, in der wir vieles einfach unkritisch hingenommen und uns allzu sehr darauf verlassen haben, dass das, was in der Zeitung steht oder im Fernsehen gesendet wird, stimmt. Meine Eltern sind in Bulgarien aufgewachsen, wo man viel stärker aufgepasst und Informationen ständig hinterfragt hat, weil man sich der staatlichen Manipulation bewusst war. Man musste sich also permanent selber eine Meinung bilden. Ich habe den Eindruck, dass wir dies hier etwas verlernt haben.

Dieser Umstand wird jetzt von manchen ausgenutzt, wodurch die Spaltung in der Gesellschaft verstärkt wird. Die Werkzeuge der künstlichen Intelligenz können nun diese Spaltung weiter vorantreiben, aber sie können auch dazu verwendet werden, um die Gesellschaft wieder zusammenzubringen. Dies setzt allerdings einen anderen Umgang mit Informationen voraus. Chat-Systeme geben einem nicht selten Antworten, die man, wenn man sie nicht hinterfragt, als korrekt ansehen kann. Erhält man also eine Information, die unplausibel klingt, sollte man mithilfe von KI unbedingt zurückfragen: „Bist du dir sicher?“ oder: „Welche Alternativen gibt es?“ Es ist also entscheidend, dass man nicht einfach die erstbeste Antwort für bare Münze nimmt, da dies auch in die Irre führen kann, sondern dass man sich umfassend informiert, um zu einer eigenen Meinung zu kommen. Auf diese Weise kann man sich mit KI sehr schnell zu neuen Themen Wissen aneignen. Hierin sind die Modelle richtig gut – insbesondere in Situationen, in denen man gar nicht weiß, wo man anfangen soll. Aufbereiten und ein gezieltes Vertiefen von bestimmten Bereichen kann man dann später. Dadurch kann man sehr viel lernen. Aus all den genannten Gründen halte ich das Bestreben von Google, Relevanz und Diversität der Resultate – also verschiedene Perspektiven – zu fördern, für wichtig und richtig.

*B.S.: D.h. also, anwendungstechnische Schulungen werden ein wesentlicher Aspekt sein?*

**S.P.:** Genau, wenngleich wir versuchen, die Applikationen derart zu gestalten, dass möglichst keine Schulung notwendig ist. Ich glaube aber, dass man trotzdem nicht darum herumkommen wird und dass der richtige Umgang mit künstlicher Intelligenz auch in der Schule unterrichtet werden sollte.

*B.S./W.W.: Im Namen der Redaktion der Dahlemer Blätter danken wir Ihnen für dieses Gespräch und wünschen Ihnen privat wie beruflich viel Glück und Erfolg.*

(Das Gespräch fand am 01. März 2025 in Berlin statt. Letzte Korrekturen wurden am 04.05.2025 vorgenommen)