



Der unbestechliche Wahrheits-Algorithmus

Wäre es nicht wunderschön, gäbe es auf dieser Erde jemanden, oder etwas, dem es gelingt, in jeder strittigen Frage schnell und ohne Restzweifel, die Wahrheit zu erkennen?

Die KI-Forscher des Neuköllner Instituts für konstruktives Denken (NIFKOD) sind der Lösung nahe und haben den Prototypen eines Algorithmus vorgestellt, der den Anforderungen bereits weitgehend entspricht. Noch bestehende „Kinderkrankheiten“, die im Testbetrieb mit ausgewählten Nutzern aufgefallen sind, sollen bis Jahresende behoben sein. Dann soll das Programm für 99 Euro als App zum Download für alle Plattformen erhältlich sein, allerdings wird es zunächst nur an lupenreine Demokraten abgegeben, um möglichen Missbrauch zu verhindern.

Wir haben uns mit dem Chef-Entwickler, Herrn Linus Gutwahr, über das Projekt und die Strategien der Software bei der Wahrheitsfindung unterhalten.

EWK: Herr Gutwahr, seit der erste Philosoph vom Himmel gefallen ist, gilt „die Wahrheit“ als das größte, sich seiner endgültigen Feststellung stets entziehende Abstraktum der Geisteswissenschaften. Wie kommt man auf die wahnwitzige Idee, etwas so Ungreifbares wie die Wahrheit mittels einer App von nicht mehr als 1,5 Megabyte unfehlbar dingfest zu machen?

Gutwahr: Nun, da muss ich mich in aller Bescheidenheit zurücknehmen und zugeben, dass die Idee dazu nicht in unserem Institut entstanden ist. Es war der damalige Justizminister, Heiko Maas, der in der Arbeit am Netzwerkdurchsetzungsgesetz nach einem einfachen Instrument suchte, das den Vollstreckern der Löschungsersuchen in den sozialen Netzwerken Rechtssicherheit verschaffen könnte. Wir wollten uns ja eigentlich weigern, weil wir den Ansatz für vollkommen unrealisierbar hielten, doch dann haben wir halt doch lieber die fünf Millionen genommen, die für die erste Studienphase ausgeworfen wurden, und uns an die Arbeit gemacht. Und das hat sich gelohnt.

EWK: Es ist Ihnen also mit Ihrem Algorithmus so etwas gelungen, wie das Ei des Kolumbus, die Quadratur der Kreises oder die Umkehr des Zeitpfeils?

Gutwahr: Bitte, bitte – übertreiben Sie nicht. Was wir gemacht haben, ist eine ganz und gar bodenständige Arbeit. Geschaffen haben wir eine App, die einfach nur feststellt, ob etwas wahr oder unwahr ist.

EWK: Das ist mir immer noch unvorstellbar. Die Welt ist voller unbekannter und nur halbverstandener Fakten, sie ist voller Meinungen, Auffassungen, Annahmen, Theorien, Hypothesen, Ideologien und Glaubenslehren, so dass die Annahme, es gäbe mindestens 10 Milliarden Fragen, die ihr Algorithmus beantworten können müsste, eher noch zu niedrig gegriffen ist. Auf welche Datenbasis stützen Sie sich dabei?

Gutwahr: Wir verwenden keine Datenbasis. Wir brauchen sie nicht. Wir sind sogar sicher, dass der Rückgriff auf bestehende Daten lediglich die Fehlerquote erhöhen würde. Niemand kann doch wissen, ob die Daten einer Datenbasis wirklich wahr sind. Nein, wir gehen einen ganz anderen Weg. Wir nennen unser Prinzip das „Prinzip der induktiven, heuristischen und prävalent-assoziativen Determinanz“.

EWK: ... ich verstehe nur Bahnhof ...

L.Gutjahr: Lassen Sie mich das Prinzip von hinter her erklären, von der Determinanz her. Nach der Lehre des Determinismus ist im Bereich des Makrokosmos vom Urknall bis zum erneuten Zusammenbruch des Universums alles kausal in Ursache(n)-Wirkungs-Ketten miteinander verknüpft. Das gilt selbst für die Kreativität, für Ideen, Erfindungen und ganze Theoriegebäude. Wo also Ursachen existieren und nachgewiesen werden können, müssen auch Wirkungen vorhanden sein.

Wir analysieren eine Wahrheitsanfrage also zunächst dahingehend, ob sie sich primär auf Ursachen oder primär auf Wirkungen bezieht. Das ist ganz einfach. Aussagen, die sich auf eine zukünftige Wahrheit beziehen, sind wirkungsorientiert, Aussagen, die sich auf gegenwärtige oder vergangene Wahrheiten beziehen, sind ursachenorientiert. Lassen Sie mich das an einem einfachen Beispiel erläutern:

Die Aussage: „Die SPD ist am Ende“, ist eindeutig ursachenorientiert, weil die Aussage ohne manifeste Ursachen nicht getroffen werden könnte. Hingegen erfordert die Aussage: „Die SPD wird 2021 noch einmal in den Bundestag einziehen“, eine wirkungsorientierte Analyse dahingehend, welche Ursachen erforderlich wären, um die Wirkung herbeizuführen. Diese Unterscheidung nach Vergangenheit, Gegenwart und Zukunft kann heute jedes einigermaßen brauchbare Übersetzungsprogramm vornehmen und wir haben uns auch eines entsprechenden Software-Moduls bedient.

EWK: Ich verstehe. Ihr Algorithmus wird dann also versuchen, entweder zu den Wirkungen passende Ursachen, oder zu den Ursachen passende Wirkungen zu finden. Aber das muss doch unheimlich schwer sein, immer in Anbetracht von mindestens 10 Milliarden möglicher Entscheidungsfälle.

Gutwahr: Nein, überhaupt nicht. Sie müssen berücksichtigen, dass wir die zu überprüfende Aussage inhaltlich in keiner Weise interpretieren. Wollten wir uns darauf einlassen, wäre eine Entscheidung nach „wahr“ oder „falsch“ doch überhaupt nicht möglich. Das wäre der Weg des Scheiterns, wie er seit Jahrhunderten, wenn nicht gar seit Jahrtausenden unentwegt gegangen wurde – mit den bekannten Ergebnissen: Mord und Totschlag! Mord und Totschlag.

Wir stellen nur fest, ob eine Wirkung bereits festgestellt ist, dann gibt das einen Wertungspunkt für „wahr“, weil „das Existente“ halt nur „wahr“ sein kann. Wird die Wirkung erst noch erwartet, gibt das einen halben Wertungspunkt für „falsch“, weil „irren“ eben nur allzu menschlich ist.

EWK: Aber, wenn Ihr Algorithmus die Aussage inhaltlich überhaupt nicht zur Kenntnis nimmt – dann ist doch eine Beurteilung vollkommen unmöglich!

Gutwahr: Sagen Sie das nicht. Es funktioniert ja, wie ich Ihnen abschließend demonstrieren werde. Allerdings habe ich noch nicht erwähnt, dass unser Algorithmus, wie jede KI, erst mit entsprechendem Übungsmaterial trainiert werden muss, und das beginnt bei jedem Anwender mit der ersten realen Aufgabe, deren Ergebnis jedoch trotzdem schon mit einer Wahrscheinlichkeit von fünfzig Prozent korrekt sein wird.

Natürlich werden alle jemals von einem bestimmten Anwender vorgelegten Abfragen gespeichert, wie auch das dazu ermittelte Ergebnis und die mimische Reaktion des Anwenders – selbst verständlich bekommt unsere App Zugriff auf die Kamera – gespeichert werden. Aus der Mimik lesen wir ab, ob der User die Wahrheit annimmt, oder ablehnt, und verwenden dieses Wissen bei den weiteren Abfragen, so dass die Wahrheit mit der Zeit immer glaubhafter wird.

Aber das ist ja noch längst nicht alles.

Jede zusätzliche neue Entscheidungsvorlage wird daraufhin überprüft, ob in der Vergangenheit bestimmte identische Buchstabenfolgen von mindestens sechs Zeichen Länge bereits einmal vorgekommen sind. Nehmen wir das bereits genutzte Beispiel: Die Buchstabenfolge „Die SPD“, mit ihren sieben Zeichen, kommt beide Male vor. Das ist der assoziative Teil, aus welchem die Prävalenz ermittelt wird. Das heißt: Die Anzahl der bereits vorgefundenen, identischen Buchstabenkombinationen bei gleicher Orientierung, also entweder in beiden Fällen ursachen- oder wirkungsorientiert, wird in ihre „wahr“ und „falsch“

Komponenten zerlegt und diese zueinander ins Verhältnis gesetzt. Je nachdem, was überwiegt, ergibt sich daraus in der aktuellen Anfrage eine Prävalenz für „wahr“ oder „falsch“.

EWK: ... und das funktioniert? Selbst dann, wenn es ganz verrückte übereinstimmende Buchstabenkombinationen gibt, die aus völlig unterschiedlichen Kontexten stammen, wie zum Beispiel – wie zum Beispiel ... Ach, jetzt fällt mir gerade nichts ein.

Gutwahr: Da machen Sie sich mal keine Sorgen. Zufällig in unterschiedlichem Kontext, noch dazu als klare Aussage formuliert, werden sie kaum jemals übereinstimmende Buchstaben- und Zeichenkombinationen finden. Da müssten Sie schon die Leerzeichen zwischen den Wörtern ignorieren, aber das tun wir nicht. Wo bei einem Nutzer gleiche Buchstabenkombinationen mit der Mindestlänge von sechs Zeichen vorkommen, geht es um die gleiche Sache. Das ist besser und umfangreicher getestet als jeder schon zum Einsatz an Menschen freigegebene Corona-Impfstoff.

Außerdem müssen Sie bedenken, dass der Algorithmus quasi in einer persönlichen Beziehung zu seinem Nutzer steht. Dessen Themenschwerpunkte – sollte er überhaupt mehrere haben, was selten genug vorkommt – verändern sich ja nicht ständig. Seit über zwei Jahren arbeitet eine bekannte Stiftung, deren Name hier nicht genannt werden soll, mit einer Beta-Version von uns, ebenso wie die meisten öffentlich auftretenden Faktenchecker.

Im Gegenzug erhalten wir von allen Wahrheitsabfragen die Log-Dateien, um die App weiter optimieren zu können. Bei der angesprochenen Stiftung zum Beispiel ist bei allen Nutzern die häufigste wiederkehrende Buchstabenfolge die siebenstellige Folge: „ist ein“. Die stammt aus Sätzen, wie „... ist ein Nazi“, „... ist ein Faschist“, „... ist ein Rechtsextremist“ – und über die assoziative Prävalenz hat sich die Antwort „wahr“ bei dieser Buchstabenfolge inzwischen zu 99 Prozent durchgesetzt – und die Anwender sind von der Treffsicherheit unserer App begeistert!

EWK: Aber das grenzt doch an Scharlatanerie! Sie verkaufen den Leuten doch letztlich als Wahrheit genau das, was sie hören wollen. Oder irre ich mich da?

Gutwahr: Nein, nein. So kann man das zwar sehen, aber so darf man es nicht sehen. Das Wesen der Wahrheit ist doch gerade ihre selektiv-individuelle Gültigkeit. Würden wir die nicht berücksichtigen, würde sie von keinem User die begehrte Fünf-Sterne-Wertung erhalten.

Aber es ist ja durchaus auch nicht so, dass jeder Anwender stets ein „wahr“ zurückerhält. Es gibt bei den erwähnten Anwendern in der Stiftung und unter den

Faktencheckern eine weitere, immer noch häufig vorkommende, sich wiederholende Buchstabenkombination, nämlich: „ist kein“. Da hat sich nach kurzer Zeit das Ergebnis „falsch“ durchgesetzt. Auch davon sind die Anwender begeistert. Was wollen Sie mehr?

EWK: Ich dachte, es ginge Ihnen tatsächlich irgendwie um Wahrheit.

Gutwahr: Selbstverständlich geht es uns um die Wahrheit, die ganze Wahrheit, die reine Wahrheit und nichts als die Wahrheit. Daher verwenden wir zur Kontrolle die so genannte induktiv-heuristische Schleife. Wir gehen vom Detail aus – und schließen auf das Ganze. Wir dividieren dabei die Summe der Buchstaben im längsten Wort der Aussage durch die Anzahl der Wörter der Aussage. Ist das Ergebnis kleiner als 1, wird der bisher gefundene Wahrheitswert invertiert, aus „wahr“ wird also „falsch“, und umgekehrt. Ist das Ergebnis größer als 1 bleibt der bis dahin ermittelte Wahrheitswert bestehen und wird so ausgegeben.

EWK: Da könnten Sie genauso gut auch einen Zufallsgenerator einsetzen ...

Gutwahr: Falsch! Wir checken damit die Komplexität. Hohe Komplexität führte bis dahin nämlich zu vielen falsch positiven oder negativen Ergebnissen. Mit Hilfe der Ergebnisinversion liegen wir inzwischen auch bei komplexen Anfragen bei weit über 90 Prozent richtiger Aussagen.

EWK: Nun, Sie hatten mir versprochen, diese App selbst ausprobieren zu dürfen. Ich bin jetzt wirklich gespannt.

Gutwahr: Sicherlich. Das machen wir. Allerdings reicht die Zeit nicht, um die App mit Ihren persönlichen Themen und Vorlieben zu trainieren. Aber wir haben hier Kopien der bereits im Einsatz befindlichen Apps einiger Anwender zur Verfügung. Auf denen könnten Sie aufsetzen. Wen würden Sie bevorzugen: Hubertus Heil, oder Wolfgang Schäuble, oder Ursula von der Leyen, vielleicht auch einfach Ralf Stegner?

EWK: Haben Sie nicht auch zufällig eine Kopie von Angela Merkel?

Gutwahr: Sie haben Glück, ist heute frisch reingekommen. Frau Merkel bekommt während ihres Urlaubs einen Update auf die neueste Version. Einen Augenblick bitte.

Augenblick: Ich verweile.

Gutwahr: Hier ist sie. Die Merkel-Wahrheits-App verfügt bereits über Sprach-eingabe. Sprechen Sie!

EWK: Nun sind sie halt mal da.

Wahrheits-App: „wahr“

EWK: Ich habe immer schon gesagt.

Wahrheits-App: „wahr“

EWK: Nur eine gemeinsame Lösung hilft.

Wahrheits-App: „wahr“

EWK: Ich habe noch nie einen großen Fehler gemacht.

Wahrheits-App: „falsch“

EWK: Lieber Herr Gutwahr, das kann doch gar nicht wahr sein, dass die App hier auf „falsch“ erkennt. Was ist denn da los?

Gutwahr: Nun ja. Eine solche Aussage hat Frau Merkel ihrer Wahrheits-App vermutlich noch nicht vorgelegt, an dieser Stelle besteht ihrerseits ja auch keinerlei Unsicherheit. Ich nehme an, dass hier der induktiv-heuristische Komplexitäts-Check zugeschlagen hat. Acht Wörter, das längste sieben Zeichen lang, ... das gibt, ... das gibt – 0,875 – und das ist kleiner als eins und folglich ist das Ergebnis invertiert worden.

Da werden wir an Merkels-Wahrheits-App doch noch ein bisschen nachjustieren müssen. Es ist zwar unwahrscheinlich, aber sollten bei ihr wider Erwarten einmal Selbstzweifel auftreten, dann darf die App diese keinesfalls noch verstärken.

Wo wären wir denn, ohne diese Frau?

EWK: Das ist eine gute Frage. Was würde die App dazu sagen?

Gutwahr: Gar nichts. An diesem Satz ist weder etwas wahr, noch etwas falsch. Sie müssten umformulieren.

EWK: Ja, stimmt. Lassen Sie mich kurz nachdenken. Dann möchte ich aber nicht mehr Merkels-Wahrheits-App verwenden, sondern möglichst eine von der SPD. Erwähnten Sie nicht Ralf Stegner. Den hätte ich gerne.

Gutwahr: Einen Moment. - So. - Sie können jetzt sprechen.

EWK: Ohne Merkel wären wir woanders.

Wahrheits-App: „falsch“

EWK: Merkel verdanken wir alles.

Wahrheits-App: „falsch“

Gutwahr: Darf ich Ihnen einen Tipp geben? Diese App ist so übertrainiert, dass jegliche Aussage als „falsch“ bewertet wird. Mit einer Ausnahme. Es muss „Stegner“ drin vorkommen, dann ist auf einmal alles wahr.

EWK: Stegner gibt den Ton an.

Wahrheits-App: „wahr“

Gutwahr: Sehen Sie, es funktioniert.

EWK: Vielen Dank für dieses Gespräch. Dank Ihrer bahnbrechenden Entwicklung ist der Sieg der Wahrheit wohl nicht mehr aufzuhalten.

Gutwahr: Gerne.

Leider fand dieses Gespräch schon vor drei Tagen statt, als die Anti-Corona-Demonstrationen in Berlin noch nicht verboten worden waren.

Ich hätte schon gerne gewusst, wie die Wahrheits-App den Sinn oder Unsinn dieses Verbotes einschätzt. So werden wir die Wahrheit wohl auch in diesem Fall wieder erst erfahren, wenn schon alles zu spät ist.