

Unsere Stimme haben sie

Unternehmen lassen heute die Stimmen von Bewerbern analysieren – und röntgen so ihren Charakter. Wie nackt stehen wir jetzt da? Und was sind die Folgen?

Von Eva Wolfangel (Text) und Tim Lahan (Illustrationen), 09.11.2018

Moment – steht da echt, ich sei schlecht organisiert? Eine Viertelstunde habe ich kürzlich mit einem Sprachcomputer telefoniert, ein harmloses Dahinplaudern: wie mein letztes Wochenende war, wie der letzte Urlaub. Ich habe nicht über Privates gesprochen, nichts angeschnitten, was mich wütend, ratlos oder deprimiert machen könnte.

Und jetzt behauptet dieser Sprachcomputer doch glatt, ich sei sehr neugierig (8 von 9 möglichen Punkten), verträglich (8), kontaktfreudig (7), risikofreudig (6), hätte ein hohes Autonomiebedürfnis (7), würde wenig Wert auf Status (5) und Dominanz (5) legen, sei nicht besonders ausgeglichen (4) – und eben nicht besonders gut organisiert (4).

Das Beängstigende: Der Algorithmus hat in fast allem recht. Ich fühle mich ertappt. Als ich später meinem Mann von den Ergebnissen erzähle, als ich in den Tagen darauf Kolleginnen davon berichte, antworten sie übereinstimmend: «Stimmt, Eva, so bist du.»

Wie kann das sein?

Unter «Aktuelle Energiebilanz» erfahre ich, dass ich während des Gesprächs gestresst war. Ich habe darüber kein Wort verloren. Aber es stimmt. Es war die Woche vor meinem Urlaub. Etliche Deadlines hingen mir im Nacken, ich hatte keine Ahnung, wie ich alles schaffen sollte. Woher weiss der Computer das? Ich fühle mich ertappt.

Weitere Sätze über mich: «Sie reagieren sensibel auf Kritik.» Und: «Sie neigen dazu, die Arbeit über andere Lebensbereiche zu stellen.» Und: «Sie kommunizieren gerne autoritär.»

Ich bin baff. Das mit der Arbeit über den anderen Lebensbereichen, das ist gerade tatsächlich mein Thema. Ich bin ehrgeizig, mein Beruf macht mir Spass, ständig habe ich ein schlechtes Gewissen meiner Familie gegenüber. Und, ja, vielleicht reagiere ich sensibel auf Kritik. Wenn wir diskutieren, höre ich zumindest von meinem Mann recht oft: «Jetzt nimm das doch nicht persönlich.» Und auch einen Dickkopf habe ich wohl. Wenn ich etwas will, finde ich meistens Wege, es durchzusetzen – wobei, mit subtilen Methoden, denn um autoritär zu sein, fehlt mir die formale Macht. Allein bei dem Punkt «schlecht organisiert» regt sich Widerspruch in mir. Ich bin eine erfolgreiche freie Wissenschaftsautorin. Wie kann ich da schlecht organisiert sein?

Und doch: Genügt es neuerdings, eine Viertelstunde lang in einen Computer zu sprechen, damit er die Persönlichkeit eines Menschen durchleuchtet?

Ich berichte seit Jahren über künstliche Intelligenz, ich kenne ihre Stärken, vor allem aber ihre Schwächen. Dann hörte ich von dem Unternehmen Precire, das behauptet, nach einem kurzen Interview den Charakter einer Person zu durchschauen. Und dachte gleich: Noch so ein Start-up, das eine Revolution verspricht, die dann nie stattfindet.

Doch dieses Mal ist es anders. Im Laufe meiner Recherche sollte ich lernen, wie tief Algorithmen inzwischen in unsere Seele blicken können. Seither bin ich beunruhigt: Wenn diese Technologie in unserem Alltag ankommt – wie nackt stehen wir dann da?

Wie Forscher begannen, Emotionen in Stimmen zu erkennen und zu klassifizieren

«Stimmen lügen nicht», sagt Klaus Scherer, der das Zentrum für Emotionsforschung an der Uni Genf gegründet hat. Er gilt als Vater dieser Forschungsrichtung – oder, mit 75, als ihr Grossvater. Seit Mitte der 1970er-Jahre erforscht er, wie Therapeutinnen Informationen aus der Stimme herauslesen können, um Rückschlüsse zu erhalten auf den Gefühlszustand ihrer Patienten.

Schon vor über vierzig Jahren hat Scherer ein Programm geschrieben, das Stimmdateien in Bezug auf Tonhöhe und Energie untersucht, um herauszufinden, wie sich die Stimme von Depressiven während einer Therapie verändert. So konnte er erkennen, bei welchen Betroffenen die Behandlung

anschluss: Bei Frauen senkte sich dann die Stimme, «das Erregungsniveau war niedriger».

Längst wird weltweit an diesem Thema geforscht. Vor einigen Jahren zeichneten Akademiker um den Amerikaner Shrikanth Narayanan von der University of California auf, wie Eheleute beim Paartherapeuten miteinander reden – und trainierten einen Algorithmus, vorherzusagen, wie sich die Ehe wohl entwickeln würde. Das System versuchte nicht, die Inhalte zu verstehen, es suchte nicht nach Stichworten oder Bedeutungen. Es analysierte nur die Stimmen, las aus Lautstärke und Tonhöhe, aus zittrigen oder brechenden Passagen die Emotionen heraus. Und wurde parallel gefüttert mit den Daten, welche Ehe hielt und welche nicht.

Nicht lange, da sagte das System mit 80-prozentiger Genauigkeit voraus, ob ein Paar nach Ende der Therapie zusammenblieb oder nicht. Und hatte damit eine bessere Trefferquote als die Therapeutinnen, die den Paaren stundenlang gegenübergesessen hatten.

Eine gruselige Vorstellung? Keineswegs, sagt Forscher Narayanan am Telefon. «Ich bin optimistisch, was die Zukunft dieser Technologie angeht. Unsere Stimme transportiert viele Informationen über unseren psychischen Status und unsere Identität.» Warum sollte man dieses Wissen nicht nutzen, um Menschen zu helfen, sie bei ihren Entscheidungen zu beraten? Und sei es bei der Frage, ob ihre Ehe eine Zukunft hat. Menschliche Wahrnehmung sei verzerrt, «sie hat einen grossen Bias», weil wir uns ablenken lassen von Äusserlichkeiten.

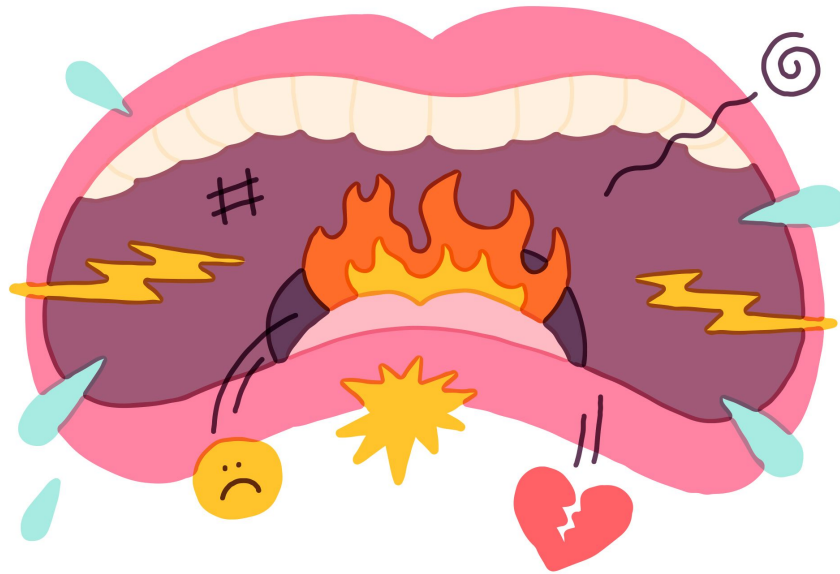
Denkt man das weiter – werden wir uns in naher Zukunft den Paartherapeuten sparen und stattdessen Amazons Alexa oder Google Home fragen, diese smarten Lautsprecher, die uns schon heute den ganzen Tag zuhören, «Hallo Google, wird meine Ehe halten?», und eine zu 80 Prozent zutreffende Aussage bekommen?

Wie die Firma Precire dieses Wissen kommerziell nutzt

Google und Amazon schweigen bislang darüber, wie sie ihre Sprachdaten auswerten. Es ist ein deutsches Unternehmen, das diese Technologie erstaunlich weit getrieben hat. Es heisst Precire Technologies, sitzt in Aachen und betreibt jenen Sprachcomputer, mit dem ich im Sommer 2018 eine Viertelstunde telefoniert habe. Mit der Erkenntnis – verzerrenderweise merke ich mir mal wieder nur die negativen Attribute –, dass ich ungeduldig, dominant und gestresst bin und Kritik persönlich nehme.

Precire analysiert Stimmdateien für Konzerne wie RWE, Talanx und Randstad. Sie alle verwenden die Software in Bewerbungsverfahren. «Wer einen Job haben will, muss den Precire-Test bei uns machen», erklärte ein Randstad-Personalchef kürzlich vollmundig.

Und ständig wollen mehr Kunden diese Software haben, für Bewerbungsverfahren und die Personalentwicklung. Warum? «Sie erfahren so mehr über einen Menschen, als sie im normalen Gespräch erfahren würden», sagt Philipp Grochowski von Precire am Telefon. Er erläutert mir meine Ergebnisse – und erträgt meine unzähligen Nachfragen mit stoischer Geduld.



«Sie reden sehr schnell», sagt er, ich sei die schnellste Sprecherin, die er je in seinen Auswertungen erlebt habe. «Das könnte Ihre Gesprächspartner überfordern», sagt er väterlich und rät dazu, eventuell langsamer zu sprechen. Er selbst wisse, wie schwer das sei: Ihm habe die Software vor Augen geführt, wie oft er «äh» und «ähm» sagt, und er gewöhne es sich gerade ab. «Okay, ich werde das mal versuchen», sage ich. Mehr, weil es sich so anfühle, als sei ich ihm das schuldig. Schliesslich hat er sich so intensiv mit meiner Art und meinem Charakter beschäftigt. Und erst im Nachhinein fällt mir auf: Das war ja gar nicht er. Das war ja nur die Maschine.

Später sehe ich auf Xing, dass er bestimmt zehn Jahre jünger ist als ich, er trägt seine dunklen Haare aufgegelt als Tolle über der Stirn, dazu Hemd und Jackett – hier wirkt er mehr wie der Unternehmensberater und nicht wie der Psychologe. Hätte ich ihn bei einer Begegnung im echten Leben anders wahrgenommen? Wohl ja. Ihm kommt seine Telefonstimme zugute.

Für mich hingegen könnte ein Gespräch im echten Leben Vorteile bringen, sagt Grochowski, denn das System hat mich erstaunlich negativ wahrgenommen, attestiert mir einen «distanzierten Kommunikationsstil», der «wenig unterstützend» ist. Das, in der Tat, erstaunt mich: Das Feedback meiner Umwelt ist exakt das Gegenteil. Meist höre ich: «Dein Lachen steckt an.» Oder: «Du hast eine wahnsinnige Begeisterungsfähigkeit.»

«Das machen Sie wahrscheinlich durch andere Kommunikationsformen wett», erklärt Grochowski: Gestik oder Mimik, Körperkontakt. Beim Telefonieren falle das weg, «das sollten Sie bedenken».

Die Software von Precire wertet zunächst aus, was gesagt wird: Welche Wörter ich verwende, was für Satzstrukturen, wie schnell ich spreche. Erst am Ende, wenn es um meine Persönlichkeit geht, spielt die Stimme eine Rolle. Ich hake nach: Was genau in meiner Stimme ist es, das auf Unorganisiertheit schliessen lässt? Grochowski weiss es nicht. Das weiss nur der Algorithmus. Wobei, Grochowski, der gelernte Psychologe, würde nie von «Unorganisiertheit» sprechen, er weiss, wie man Schwächen schönredet. «Sie sind flexibel», sagt er.

Precire hat seinen Algorithmus trainiert, indem das Unternehmen eine künstliche Intelligenz die Sprachdaten von 5000 Probanden hat auswerten lassen – und diese Menschen zudem umfangreich psychologisch vermass. Und nun sucht die Software nach Mustern: Wo haben diese 5000 Pro-

banden Gemeinsamkeiten mit mir? Was zeichnet jene Probanden aus, die ähnlich sprechen wie ich?

Wobei kein Mensch weiss, was dieses «ähnlich» ist: Es geht um Hunderte von Faktoren in meiner Stimme, kein Ohr würde darin Muster finden. Es ist viel zu komplex. Die Algorithmen hingegen sind fabelhafte Mustererkenner. Auf diese Weise kommt die KI zu Schlüssen über meine Persönlichkeit, ohne die entsprechenden Informationen zu haben. Es sind Erkenntnisse, für die Psychologen umfangreiche Tests machen müssten.

Aber: Wie zuverlässig sind diese Muster? Sagen solche Korrelationen wirklich etwas aus?

Nur ein Beispiel: Forscher haben einmal festgestellt, dass Menschen, die Filzgleiter für ihre Stühle kaufen, um das Parkett zu schonen, Kredite besonders zuverlässig zurückzahlen. Warum das so ist? Niemand weiss es. Und es kann den Banken ja auch egal sein. Es gibt diese Korrelation, das Kaufen von Filzgleitern ist eine unverfängliche Information, man kann sie womöglich bei Amazon oder Google oder sonstwo einkaufen – und erhält eine Aussage über die Kreditwürdigkeit eines Menschen.

Ist es bei Stimme und Emotionen auch so? Nach dem Motto: Ganz gleich, was genau die Kausalitäten sind, Hauptsache, das Ergebnis stimmt?

Wie komplex es ist, Emotionen in Stimme zu analysieren

Ganz so einfach ist es nicht, findet der Genfer Forscher Klaus Scherer. Er will nicht nur wissen, wieso psychologische Faktoren Sprache verändern, er will Emotionen obendrein sehr viel differenzierter entschlüsseln. Wenn er davon spricht, dass unsere Stimme Emotionen vielschichtiger abbildet als unser Gesicht, meint er auch: Emotionen sind komplexer, als viele Modelle der Psychologie annehmen.

Das klassische Emotionsmodell von Paul Ekman aus den 1970er-Jahren unterscheidet nach sieben Grundemotionen, die in allen Kulturen zu finden sind und die sich vor allem in der Mimik ausdrücken: Angst, Wut, Ekel, Freude, Trauer, Verachtung und Überraschung.

Doch dieses Modell helfe da nicht weiter, sagt Scherer: «In unseren Stichproben sind 80 bis 90 Prozent aller Emotionen eine Mischung aus verschiedenen Grundemotionen», beispielsweise Trauer und Wut. Um diese zu erkennen, dafür ist die Stimme wiederum gut: «Die Stimme ist komplex und kann Mischungen besser abbilden als das Gesicht», sagt Scherer.

Doch viele dieser Mischungen kommen nicht allzu häufig vor. Deep-Learning-Verfahren, jene tiefen neuronalen Netze, die besonders gute Mustererkenner sind, brauchen riesige Mengen an Trainingsdaten. Obendrein sind gerade Emotionsdaten schwierig zu bekommen, klagt Scherer: «Sie sind zu privat.» Deshalb arbeitet er mit seinen Kollegen unter anderem mit Schauspieler-Stimmen – auch wenn diese «nur so tun als ob», gehören sie offenbar zu den wenigen Menschen, die «mit der Stimme lügen» können: Es gehört zu ihrem Beruf.

Wie trainiert man einen Algorithmus, Emotionen zu lesen?

Der Informatiker Björn Schuller von der Universität Augsburg will die Emotionserkennung aus der Stimme nutzen, um psychisch kranken oder geistig behinderten Menschen zu helfen – und ihren Angehörigen. Er hat das Unternehmen Audeering gegründet und eine App entwickelt, die Emotionen von autistischen Kindern erkennen kann, ebenfalls anhand ihrer Stimme. Sein System sagt schon bei kleinen Kindern erfolgreich vorher, ob diese später die Diagnose Autismus bekommen. Auch er erforscht seit Langem, wie Stimme und Emotionen zusammenhängen.

Er ist ein mitreissender Redner. Wer dem ehrgeizigen 43-jährigen Informatiker mit der blonden Sturmfrisur auf einer Konferenz zuhört, wird immer wieder lachen über seine Pointen. Er begeistert. Später verrät er mir, dass er einen Trick anwendet: Er hat vor einigen Jahren die Videos der erfolgreichsten Youtuber ausgewertet – mit ebendiesem System, das Emotionen allein aus der Stimme erkennt. Schliesslich konnte sein Algorithmus vorhersagen, wie beliebt ein Youtube-Film sein wird. Der Algorithmus habe ein zentrales Muster gefunden: Die Abwechslung zwischen positiven und negativen Emotionen. Folgte eine Show diesem Muster, war sie erfolgreich.

Also macht er es nun auch so in seinen Vorträgen: Zwischen zwei Highlights platziert er einen nachdenklichen Inhalt, eine Herausforderung, eine ungelöste Frage. Seine Beliebtheit explodierte.

Heute, nach vielen Jahren des Experimentierens, hat er drei zentrale Faktoren ausgemacht, um Stimmen zu klassifizieren: erstens Erregung, also wie stürmisch oder begeistert jemand spricht. Zweitens Valenz: Ist die Grundstimmung eher positiv oder negativ, schwingt Ärger oder Furcht mit oder Freude oder Glück? Und drittens Dominanz: wie laut jemand spricht.

Kombiniert man diese drei Kategorien, «so kann man einen hochdimensionalen Emotionsraum aufspannen», sagt Schuller: So könne eine hohe Erregung allein beispielsweise sowohl Freude als auch Ärger bedeuten. Kombiniert man die Erregung in der Stimme aber mit der Valenz, der Grundstimmung, kann man Ärger und Freude unterscheiden. Mittels Dominanz könne man wiederum zwischen Ärger und Furcht unterscheiden.

Auch er ringt damit, genügend Daten für seine Modelle zusammenzubekommen. Im Fall der autistischen Kinder braucht Schuller bis zu sechzig übereinstimmende Angaben von Eltern oder Betreuern darüber, ob beispielsweise ein Kind gerade glücklich klingt. Und dann ist das nur ein Beispiel, aus dem das System lernen kann. Die modernen Verfahren maschinellen Lernens brauchen extrem viele solcher Beispiele als Trainingsdaten – im Idealfall sind es Tausende.

Auch der Augsburger Professor kann seine Technologie nicht nur «für das Gute» einsetzen, auch er braucht kommerzielle Anwendungen, um seine Forschung zu finanzieren. Die Kunden seines Start-ups Audeering sind beispielsweise Marktforschungsunternehmen, die nicht nur am Inhalt, sondern auch an der Stimme der Befragten hören wollen, wie diese ein Produkt bewerten. Schliesslich sind solche Bewertungen ehrlicher – weil Stimmen nicht lügen. Und auch die Analyse von Stimmen aus dem Internet (beispielsweise von Youtube-Filmen), die über bestimmte Produkte sprechen, sei ein riesiger Markt: «Da können Sie Meinungsbildung im Netz in Echtzeit verfolgen.»

Schuller ist überzeugt, dass man schon bald «richtige Gespräche» mit Computern führen kann – anstatt der aktuellen Geräte, die man zwar per Sprache steuern kann, die sich aber emotional nicht an ihr Gegenüber anpassen. «Im menschlichen Gespräch muss der Dialog weiterlaufen, und dafür brauche ich die Emotionen», sagt Schuller. Geräte der Zukunft reagieren emotional passend, was auch für das Verkaufen der Zukunft zentral ist: «Der Computer kann dann perfekt die Stimmung analysieren, und weiss, wann ich schwach bin.» Aber wollen wir das? Wollen wir, dass Maschinen wissen, wann wir schwach sind? Dass sie es spüren, besser, als unsere Nächsten es spüren? Und das alles aus Daten, die wir nicht schützen können, weil wir sonst schweigen müssten? Und was ist mit den Fehlentscheidungen?

Was ich über mich gelernt habe

Eines Abends sitze ich mit zwei Freundinnen zusammen, beide sind Ingenieurinnen, sie reden auf mich ein. Ich würde brauchen: Aufmunterung. Ich bin überlastet, habe zu viele Aufträge angenommen und die dunkle Ahnung, dass es diesmal wirklich eng wird. Und nun purzeln aus ihren Mündern Begriffe wie Excel, Planungstool, Projektmanagement. Sie fragen: «Wie lange brauchst du für einen Artikel?» Darauf ich: «Keine Ahnung, ganz verschieden.» Sie rollen mit den Augen und sagen: «Wir bauen dir da mal ein Programm.»

Daraus wurde nichts, weil ich eine natürliche Abneigung gegen Planung habe. Geplantes Texteschreiben? Wie soll das funktionieren?

Mir schwant: Precire hat auch in diesem Punkt recht. Wenn man es so sieht, bin ich nicht besonders gut organisiert. Ich erledige meine Arbeit mehr nach Gefühl. Was dringend scheint, wird zuerst gemacht. Und auch mein Überblick, ob ich gerade viel zu tun habe oder schon zu viel, ist nicht objektiv, sondern gefühlt.

Meistens geht das gut, vermutlich auch wegen meiner von der Precire-KI attestierten «hohen Verausgabungsbereitschaft» (8 von 9). Ich habe noch keine Deadline gerissen. Ausser die des Finanzamtes. Die Steuererklärung fühlt sich irgendwie erst dann dringend an, wenn die Frist schon abgelaufen ist. Da kommt mir zum Glück meine ausgeprägte Fähigkeit zu Hilfe, «steuernd zu kommunizieren», die mir Precire bescheinigt, und meine «natürliche Überzeugungskraft»: ein Anruf beim Amt, und die Frist ist verlängert.

Werden wir uns in Zukunft besser verstehen?

Wie wahrscheinlich ist es, dass diese Technologie zum Schaden von uns Menschen eingesetzt wird, um uns zu kontrollieren und unsere Freiheit zu beschneiden? Beide Forscher, Björn Schuller und Klaus Scherer, wollen Kranken helfen. Welche Büchse der Pandora öffnen sie mit ihren Erkenntnissen?

Klaus Scherer, der Forscher aus Genf, macht sich Sorgen. «Künstliche Intelligenz wird dank der Stimme in Zukunft mit Sicherheit Emotionen gut erkennen können, und das bringt die Gefahr mit sich, dass ohne unser Wissen Informationen über unsere emotionalen Reaktionen gespeichert werden.» Angesichts des Skandals um Cambridge Analytica, das Facebook-Daten nutzte, um die politische Gesinnung von Menschen vorherzusagen, sieht er keinen Grund, Vertrauen in die amerikanischen Konzerne zu haben. «Amazons Sprachbox Alexa speichert schon jetzt alle Daten auf Servern in den USA. Im Augenblick sehe ich die Gefahren höher als den Nutzen dieser Technologie.»

Björn Schuller in Augsburg pflichtet dem bei. «Aus unserer Stimme kann man so reichhaltige Informationen bekommen, das reicht vom Alter über die Gesundheit bis hin zu Gefühlen. Für totalitäre Systeme ist das eine gute Quelle, um Überwachung zu perfektionieren.»

Shrikanth Narayanan, der Amerikaner, dessen Rechner das Scheitern von Ehen vorhersagte, widerspricht. «Dank solcher Tools könnten wir genau erfahren, wie es unserem Gegenüber geht.» Das könnte beispielsweise der Callcenter-Kommunikation helfen: Die Mitarbeiter würden immer genau richtig auf die Emotionen des Anrufers reagieren. Oder in der Kinderklinik, so Narayanans Vision: «Dem Pflegepersonal fehlt oft die Zeit, sich in die Kinder hineinzusetzen.» Wie viel besser wäre es, wenn ihnen die Maschine einen Einblick in die Gefühlslage eines Patienten gäbe. Seelenlesen, effizient.

Wie viele Begegnungen, schwärmt der Forscher, könnte die Technik auf eine höhere Stufe heben, wenn wir den psychischen Zustand unseres Gesprächspartners kennen würden. «Wir könnten uns auf einer viel tieferen Ebene verstehen!» Natürlich müsse man Grenzen ziehen. «Aber die Technologie ist da, wir können sie nicht verhindern, wir sollten sie für das Gute verwenden.»

Doch was ist das Gute? Was ist die Folge dieser übermenschlichen Kommunikation, wenn wir mit einem Klick oder einem Blick auf ein Display wissen, wie es im Herzen unseres Gegenübers aussieht? Sollen wir unseren Partner lieber gleich verlassen, wenn der Algorithmus eine Scheidung in drei Jahren vorhersagt? Wozu Zeit vergeuden? Werden wir mögliche Freunde künftig scannen nach ihren inneren Werten, ehe sich diese in ihren Handlungen, im freundschaftlichen Miteinander offenbaren?

Nein. Wir werden dann erkennen, warum wir so sind, wie wir sind. Warum wir eben nicht – wie der Precise-Computer – nach wenigen Minuten Gespräch die Persönlichkeit eines Menschen in einer Tabelle auflisten können. Unser Unvermögen hat einen Sinn, nämlich den, dass wir anderen Menschen eine Chance geben, anstatt unsere Freundschaften mittels Algorithmen zu optimieren. Dass wir offen sind füreinander. Weil zwischenmenschlicher Kontakt viele Facetten hat und sich nicht berechnen lässt.

Zur Autorin

Eva Wolfangel ist Wissenschafts- und Reportagejournalistin, Moderatorin und Speakerin. Sie veröffentlicht unter anderem in der Zeit, der Süddeutschen, Geo und Spiegel. Im vergangenen Juli wurde sie mit dem European Science Writer Of The Year Award ausgezeichnet.