

BASISARTIKEL

Künstliche Intelligenz im Religionsunterricht:

Wer soll den Menschen richten dürfen?

Von Prof. Dr. Katharina Zweig

Künstliche Intelligenz ist ein Schlagwort, dem man kaum entgehen kann. Aber was genau ist damit gemeint? Das ist gar nicht so einfach zu sagen, denn manche interpretieren es als das Forschungsfeld, andere als die Menge der Methoden, die dieses Feld entwickelt hat, und wiederum andere meinen damit die Software, die Computern „Intelligenz“ verleiht. Im Wesentlichen geht es aber bei allen drei Interpretationen um die Frage, wie Computer solche Arbeiten ausführen können, die – wenn sie von Menschen ausgeführt werden – von diesen einige Intelligenz abverlangen. Diese Definition führt natürlich nicht viel weiter, weil die Wissenschaft auch nicht viel besser definieren kann, was denn nun die menschliche Intelligenz ausmacht. Aber die meisten von Ihnen erkennen sie doch, wenn Sie sie sehen.

Die genannte Definition ist auch relativ breit: Selbst ein schnöder Taschenrechner kann mathematische Berechnungen durchführen, die für die meisten von uns Intelligenz erfordern, und doch würde ihn wohl niemand ernsthaft der „künstlichen Intelligenz“ zurechnen. Tatsächlich ist es so, dass der Begriff künstliche Intelligenz ein bewegliches Ziel zu sein scheint. Während wir in den 1980er Jahren noch davon überzeugt waren, dass nur Menschen im Schach brillieren können, und maschinell übersetzte Texte eher zu Lachanfällen führten (siehe Abbildung 1), zeigten die letzten Jahre, dass Computer Autos lenken und selbst das japanische Brettspiel Go meistern können.

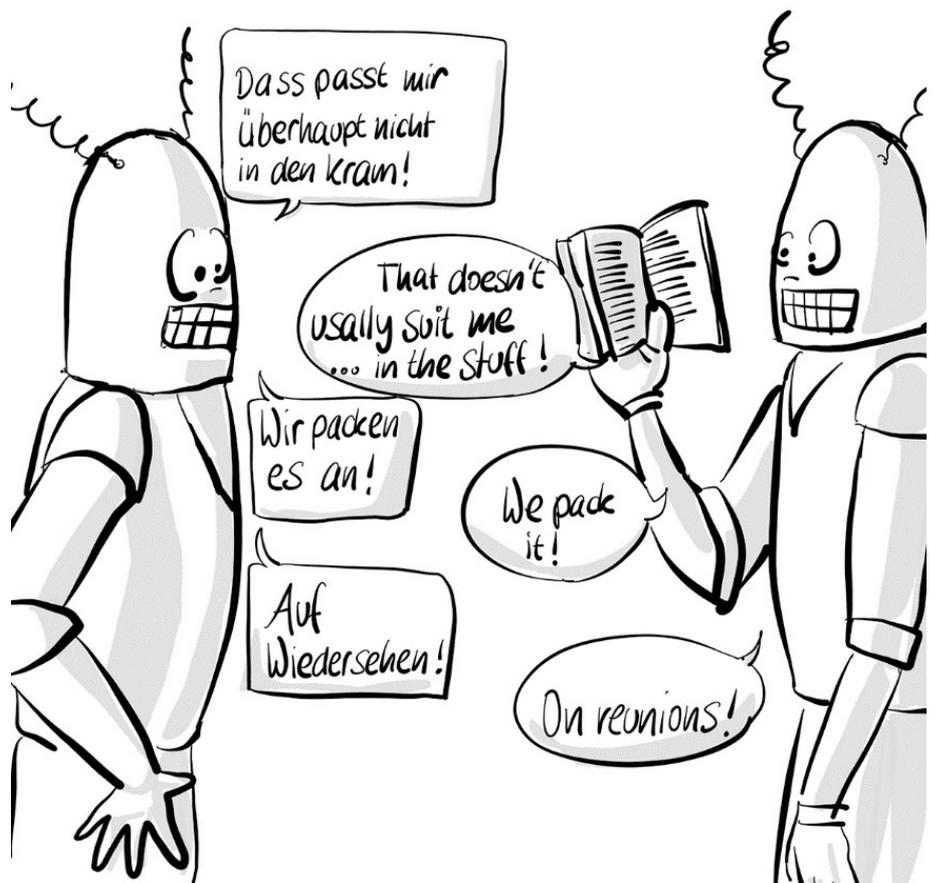


Abbildung 1: Übersetzungsversuche eines maschinellen Übersetzungssystem aus den 1980ern. ((c) Sandra Schulze, mit Genehmigung aus dem Buch „Ein Algorithmus hat kein Taktgefühl“ der Autorin).

Tatsächlich liegt das daran, dass sich die Art und Weise geändert hat, mit der Maschinen Sinn und Ordnung in der Außenwelt erkennen. In den 1980er Jahren wurden den Computern manuell zusammengestellte Informationen und Entscheidungsregeln von Expertinnen und Experten vorgelegt. Mit Informationen wie: „Kaninchen sind Säugetiere“ und „Säugetiere sind Tiere“ konnten die Com-

puter ableiten, dass auch Kaninchen Tiere sind. Das klingt in diesem Beispiel nicht weltbewegend, hilft aber dabei, Informationen kompakt zu speichern und bei Bedarf „auszurechnen“. Solche sogenannten Ontologien und die von Experten in Expertensystemen computergerecht zusammengefasste Entscheidungsregeln dominierten die erste Forschungsphase zu künstlicher Intelligenz.

Mit solchen Expertensystemen in Form von riesigen Grammatiken machten Informatiker dann auch ihre ersten Versuche in der maschinellen Übersetzung – mit eher weniger Erfolg.

In den letzten Jahrzehnten aber trat ein Paradigmenwechsel ein: Anstatt sich auf menschengemachte Entscheidungsregeln oder von Menschen kodierte Informationen zu verlassen, wurden Methoden entwickelt, um in großen Datenmengen nach Auffälligkeiten suchen zu lassen. Für die maschinelle Übersetzung funktioniert das so: Man nimmt eine große Menge von Texten, die es sowohl in der einen als auch in der anderen Sprache gibt. Dafür, dass es solche Texte zu Hauf gibt, hat unter anderem die Europäische Union gesorgt, die routinemäßig alle Texte in die Sprachen aller Mitgliedsstaaten übersetzen lässt. In solchen Textzwillingen sucht nun eine Maschine mit Hilfe von Methoden des sogenannten „maschinellen Lernens“ nach Paaren von Wortgruppen, von denen die eine (fast) immer im Text der ersten Sprache auftaucht, wenn die andere im Text der zweiten Sprache auftaucht und umgekehrt. Die Maschine schließt daraus, dass die eine Wortgruppe die Übersetzung der anderen sein könnte. Anstatt also regelbasiert das eine in das andere zu übersetzen, finden statistische Auswertungen statt, die auszählen, wie oft das eine Muster mit dem anderen Muster einhergeht. Basierend auf dieser „Statistik auf Steroiden“, wie die künstliche Intelligenz immer wieder scherzhaft genannt wird, sind Übersetzungen bemerkenswert

gut geworden. Davon kann man sich leicht auf der Webseite der deutschen Firma DeepL überzeugen, indem man dort einen Text in ein Formular eingibt und es sich automatisiert übersetzen lässt.

Und nicht nur in der Übersetzung funktioniert dieser neue Ansatz ungemein gut: Mit großen Datenmengen und hoher Rechengeschwindigkeit können Computer Muster in Bildern erkennen, das Kaufverhalten von Konsumenten vorhersagen oder gesprochene in geschriebene Sprache umwandeln. Die grundlegende Methode besteht immer darin, in den Daten nach Korrelationen zu suchen, also statistisch auffälligen Gleichzeitigkeiten in den Daten. Wenn z.B. viele Personen, die den Film „Star Wars III“ mochten, auch den Film „Star Wars I“ mögen, kann man Personen, die nur einen der beiden kennen und gut bewerten, auch den zweiten Film empfehlen. Aber eine solche Methode kann sich natürlich auch irren, wie jeder, der über Aberglauben den Kopf schüttelt, schon weiß: Auch wenn schwarze Katzen gern einmal unter Leitern einhergehen und diese dann manchmal auch umfallen, muss weder das eine noch das andere zwangsläufig immer zu Unglück führen. Einem solchen Aberglauben können auch Computer verfallen: In einer meiner Forschungsarbeiten behauptete ein Algorithmus, dass Menschen, die einen Kinderfilm aus der Serie „VeggieTales“ mögen, sicherlich auch „Sex in the City“ mögen werden. Aber hier saß der Computer einem Irrtum auf: Denn vermut-

lich handelt es sich hier um Mütter, die auf der einen Seite Filme für ihre Kinder ausleihen und „ liken“ und auf der anderen Seite auch noch einen eigenen Filmgeschmack besitzen – eine Korrelation, die aber nicht dazu führen sollte, allen Frauen, die „Sex in the City“ mögen auch die Karotten- und Lauchpuppen von VeggieTales zu empfehlen.

Solche Muster zu finden, auch wenn sie nicht 100% gültig sind, sondern nur 80% der Personen betreffen, ist im Online-Handel immer noch gut geeignet, um den Profit zu erhöhen. Denn die dafür notwendige Datenverarbeitung ist so schnell und billig, dass sich eine solche Empfehlung schon dann lohnt, wenn nur einige wenige Kunden und Kundinnen dazu verführt werden können, ein weiteres Produkt zu kaufen. Aber der große Erfolg solcher Auswertungen und der daraus resultierenden Vorhersagen im Online-Handel führt zunehmend dazu, dass Firmen anbieten, auch in anderen Fragen das Verhalten der Menschen auszuwerten und ihr zukünftiges Verhalten vorherzusagen. Und hier geht es um eine Frage, die aus meiner Sicht einen Platz im Religionsunterricht haben könnte: „Wer soll den Menschen richten dürfen? Gott, Mensch, oder Maschine?“

Tatsächlich ist die Bibel ja voller Geschichten, in denen Gott Menschen bewertet – sie für gerecht erklärt, sie straft oder belohnt. Und auch die weltliche Gerichtbarkeit wird erwähnt. Wie steht es mit einer maschinellen Gerichtbarkeit? →

Wie man aus Daten solche Entscheidungsregeln ableiten kann, können Sie in Abbildung 2 selbst ausprobieren. Die Abbildung zeigt eine erfundene Datenmenge, zu der ich immer folgende Geschichte erzähle: Angeblich könne man die Neigung zur Kriminalität einer Person an der Blutkonzentrationen von zwei Hormonen mit den schönen Namen „Sanftosan“ und „Kriminolin“ erkennen – das habe eine noch geheime Forschungsarbeit an der TU Kaiserslautern ergeben. Das Datenset beschreibe Menschen anhand dieser beiden Werte und ordnet ihnen noch eine Farbe zu, die angibt, ob sie tatsächlich für eine Straftat verurteilt wurden (roter, quadratischer Smiley) oder bisher nicht verurteilt wurden (grüne, runde Smileys).

Tatsächlich ist visuell in diesem fiktiven Datenset ein gewisses Muster erkennbar: Die kriminellen Personen sind eher oben zu finden, die unschuldigen eher unten. Eine der vielen Methoden der künstlichen Intelligenz würde nun versuchen, eine Trennlinie so zwischen die Datenpunkte zu legen, dass möglichst alle roten Smileys auf der einen Seite liegen und alle grünen auf der anderen Seite. Dies können Sie nun mit einem Lineal und Stift simulieren: Zeichnen Sie dafür mit einem Stift eine gerade Linie durch die Datenpunkte, die sie für optimal halten, um die roten (quadratischen) möglichst gut von den grünen (runden) Datenpunkten zu trennen.

Diese Linie ist ab jetzt Ihre „Entscheidungsregel“, nach der Sie weitere Menschen bewerten können. Prüfen Sie den folgenden Datenpunkt: Frau Müller hat einen Sanftosan-Level von 5,5 (auf der horizontalen Achse zwischen 5 und 6) und einen Kriminolin-Level von 4 (auf der vertikalen Achse). Dieser Punkt ist in der Abbildung 2 mit X bezeichnet. Liegt dieser Punkt bei Ihnen auf der „grünen“ oder der „roten“ Seite? Die Position bezüglich der Linie bestimmt nun Ihr Urteil

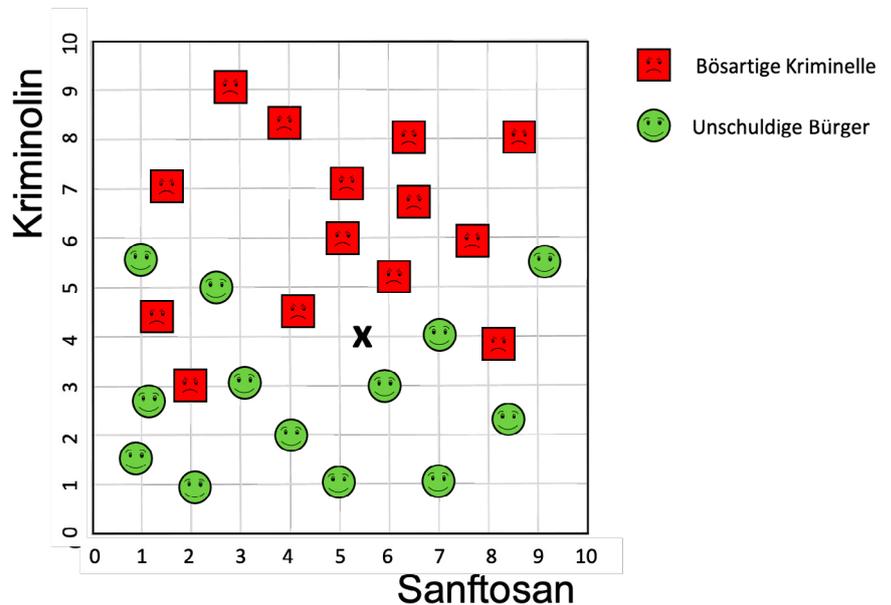


Abbildung 2: Ein fiktives Datenset. Rote, quadratische Smileys repräsentieren die Blutlevel der beiden genannten Hormone „Sanftosan“ und „Kriminolin“ von Kriminellen, die der grünen, runden Smileys von bisher nicht verurteilten Bürgerinnen und Bürgern. Das X bezeichnet einen neuen Datenpunkt.

über Frau Müller. Erfahrungsgemäß werden die meisten von Ihnen Frau Müller als „unschuldige Bürgerin“ bezeichnen, aber einige sie auch auf der kriminellen Seite sehen. Damit könnten wir jetzt eine neue Prozedur vor Gericht verwenden: Kriminellen wird einfach vor der Urteilsverkündung ein bisschen Blut abgenommen, sie werden „gewogen“ und dann eventuell – Mene tekell! – als „zu leicht“ verworfen.

Wenn Sie gewusst hätten, dass die Trennlinie zur Bewertung von Personen eingesetzt werden wird, hätten Sie vielleicht lieber eine andere Linie gewählt. Vielleicht hätten Sie dem Rechtsphilosophen William Blackstone zugestimmt, der 1760 sagte, dass lieber zehn Schuldige entkommen mögen, als dass ein Unschuldiger ins Gefängnis käme. Ganz biblisch könnten Sie auch fragen, „Willst du auch den Gerechten mit dem Ruchlosen wegraffen.“ (Gen 18,23). Diese Frage stellt Abraham Gott, als dieser die Stadt Sodom vernichten will. Doch Abraham appelliert an Gottes Gerechtigkeit: Wenn

es nur zehn Gerechte in der Stadt gebe, dann wird er der ganzen Stadt vergeben (vgl. Gen 18,16–33). Wenn Sie auch so denken, dann würden Sie Ihre Linie vielleicht weiter nach oben legen. Ginge es allerdings um etwas sehr Schwerwiegendes, wie Mord oder Terrorismus – wäre es Ihnen dann vielleicht lieber gewesen, wenn Ihre Trennlinie lieber auf der „sicheren Seite“ läge, also weiter unten?

Das fiktive Beispiel zeigt: Welchen Bewertungsmaßstab wir anlegen wollen, das hängt von der Situation ab. Den Berechnungsmethoden der Maschine ist dieser Kontext aber nicht bekannt. Sie bekommen nur die Datenpunkte und deren Verhalten, das sie nachher bei anderen Personen vorhersagen sollen. Eine Maschine würde immer dieselbe Linie legen – unabhängig davon, ob die Datenpunkte Handtaschendiebe oder Mörderinnen zeigen.

Die Entscheidungsregel, die Frau Müllers Schicksal bestimmen könnte, ist zudem abhängig von Daten, die wir von anderen

Personen aus deren Vergangenheit hatten. Die mathematischen Methoden, die die Entscheidungsregel bestimmten, haben Frau Müller und alle anderen Menschen, die bewertet werden sollen, damit Gruppen zugeordnet, denen sie angeblich gut genug ähneln, um daraus ihr zukünftiges Handeln ableiten zu können.

Ich nenne diese Methode eine algorithmisch legitimierte Vorurteilsbildung, wobei wir uns als Gesellschaft doch immer mehr bemühen, verdeckte und offene Vorurteile zu vermeiden. Und gerade im Neuen Testament finden sich so viele Geschichten, in denen Jesus die Menschen ganz anders beurteilte, als es seine Mitmenschen getan hätten. Vom Samariter, einer weithin verhassten Volksgruppe, der man nichts Gutes zutraute, berichtete er als dem „Barmherzigen Samariter“ (Lk 10,25–37). Die Ehebrecherin ließ er laufen (Joh 8,1–11), den Zöllner holte er vom Baum herunter, um mit ihm zu sprechen (Lk 19,1–10).

Das Menschenbild hinter der Verwendung von gelernten Entscheidungsregeln aus großen Datenmengen schreibt Menschen in ihrer zukünftigen Entwicklungsfähigkeit fest. Es kennt auch notwendigerweise weder Gnade noch Vergeltung: Alle, die gewisse Eigenschaften erfüllen, werden von der Maschine gleich behandelt. Das ist gut, wenn es „ohne Ansehen der Person“ geschieht. Die Erfahrung in den letzten Jahren hat aber gezeigt, dass noch nicht einmal das immer gegeben ist: Wenn in den Daten schon diskriminierende Muster enthalten sind, wird die Maschine diese in ihren Entscheidungsregeln verankern und dadurch sowohl verstetigen als auch erhärten.

Und im Gegensatz zum Menschen kann die Maschine niemals aus dem Kontext heraustreten, unter dem die Daten ursprünglich zur Verfügung gestellt wurden. Wo König Salomo in seinem Ge-

richtssaal eines Tages etwas völlig Unerwartetes tat, um herauszubekommen, wer die Mutter des noch nicht sprechfähigen Babys sei (vgl. 1 Kön 3,16–28), bleibt die Maschine immer in ihrem starren Bewertungsschema. Auch das entspricht wohl kaum dem christlichen Grundgedanken, das Individuum in den Vordergrund zu stellen.

Wir alle werden täglich bewertet – von Mensch und Maschine. Und oftmals ist die maschinelle Bewertung hilfreich – z.B. wenn sie uns genau die Webseite aus Milliarden heraus präsentiert, die wir gerade gesucht haben. In anderen Fällen ist ihr Urteil (oder Fehlurteil) höchstens ein kleines Ärgernis – etwa dann, wenn der Onlineshop meines Vertrauens mich für kleinkariert hält und mir nur noch ebensolche Oberteile präsentiert.

Der Religionsunterricht könnte fragen, wann es über dieses kleine Ärgernis hinausgeht und wo das maschinell verankerte Menschenbild unserem christlichen Menschenbild mit seiner Perspektive auf das Individuum so zuwiderläuft, das wir darauf verzichten sollten. Denn tatsächlich ist das Beispiel nicht so fiktiv: In den USA werden Menschen vor Gericht tatsächlich von Maschinen bewertet in Bezug auf ihr „Rückfälligkeitsrisiko“ – zwar nicht anhand der erfundenen Hormone, sondern anhand von Fragebögen und den Daten, die die Behörden so über sie haben. Aber soll so unsere Zukunft aussehen? Eine Bewertung in Form von starren Regeln, ohne Möglichkeit der Nachfrage, ohne zuzugestehen, dass Menschen sich ändern können, und ohne die Möglichkeit, Gnade walten zu lassen?

In Form der künstlichen Intelligenz bekommt somit eine alte Frage ein neues Gewand: Wer soll den Menschen richten?

Und es ist bei Weitem nicht die einzige Frage (nicht nur für den Religionsunter-

richt), die die Nutzung von künstlicher Intelligenz aufwirft, da es ja Methoden sind, die Probleme lösen sollen, die von einem menschlichen Problemlöser Intelligenz erforderten. Die vielen moralischen Probleme, die es dabei schon immer gab, stellen sich jetzt noch intensiver, denn die Behandlung durch den Computer verlangt, dass wir unsere menschlichen Werte ganz genau formulieren und formalisieren können – und Informatiker/innen müssen dafür ausgebildet werden, zu wissen, wo das prinzipiell technisch nicht möglich ist. Damit wir gemeinsam die Welt gestalten, in der wir leben wollen.



Katharina Zweig ist Professorin für Informatik an der TU Kaiserslautern, wo sie das Algorithm Accountability Lab und das Center for Ethics and the digital Society (CEDiS) leitet. Sie war Mitglied in der Enquete-

Kommission für Künstliche Intelligenz und hat den Bestseller „Ein Algorithmus hat kein Taktgefühl“ geschrieben.