

## Künstliche Intelligenz im Recht

### Was geht, was nicht und wo geht die Reise hin?

Dr. Paul von Büнау, Geschäftsführer, idalab

21. Mai 2018

LR 2018, Seiten 80 bis 83 (insgesamt 4 Seiten)

---

Nachdem das Interesse an KI jahrzehntelang im Winterschlaf befindlich schien, entlädt es sich im aktuellen AI-Summer in einer vormals nicht gekannten Euphorie. Wohin man auch schaut, wird prophezeit, dass tiefgreifende Veränderungen anstehen und dass, wer jetzt nicht aufspringt, den Anschluss verpassen wird. So auch in den Rechtswissenschaften, auch wenn die KI-Welle hier vergleichsweise spät kommt. Hier hält die Künstliche Intelligenz als Teil des Legal Technology Booms Einzug und beginnt bereits langsam, ihr disruptives Potenzial zu entfalten. Doch während jenes Potenzial durchaus ernst zu nehmen ist und in der Tat unbestreitbar Veränderungen anstehen, sollte der weit verbreiteten Angst, Künstliche Intelligenz könnte den Juristen zuerst bevormunden und dann völlig abschaffen, mit Vorsicht begegnet werden. Stattdessen tut man gut daran, der Frage „Was ist Künstliche Intelligenz?“ ganz dezidiert nachzugehen – um die aktuellen Entwicklungen kritisch beurteilen zu können.<sup>1</sup>

1

#### 1. Künstliche Intelligenz: Eine Kombination aus Algorithmen und Daten

In der Fachliteratur wird es oft als Abgrenzungsmerkmal zwischen künstlich intelligenten Systemen und „normaler Software“ betrachtet, wenn ein System eine Machine Learning Komponente enthält. Ist dies der Fall, so sagt man, dass das System implizit programmiert wird. Nahezu alle IT-Systeme sind explizit programmiert, Beispiele hierfür sind Taschenrechner, oder eine ERP-Software wie SAP. Implizite Programmierung bedeutet, dass die Schritte zur Lösung einer Aufgabe nicht von vornherein bekannt sind. In einem implizit programmierten System „lernt“ ein Algorithmus aus einem historischen Datensatz ein erwünschtes Verhalten. Das heißt, er erkennt Muster und Zusammenhänge zwischen den Eingangs- und Ausgangsgrößen und übersetzt dieses „Erlernte“ in Systemverhalten. Im Kern eines jeden intelligenten Systems stehen also Algorithmen und Daten. Der Algorithmus ist zu verstehen als in einzelne Schritte heruntergebrochene Anweisungen, die der Computer maschinell abarbeiten kann,

2

---

<sup>1</sup> Dieser Artikel von Autor Dr. Paul von Büнау ist bereits in der Studie „Auf dem Weg zur Legal Technology 3.0“ (<http://unternehmensjuristen.wolterskluwer.de>), die von Wolters Kluwer veröffentlicht wurde, erschienen.

während die Daten, ohne die der Algorithmus wertlos ist, den unverzichtbaren Baustoff des Systems darstellen.

Die Wichtigkeit der Daten für ein künstlich intelligentes System kann gar nicht genügend betont werden. Von ihrer Qualität und Quantität hängt heute viel mehr ab als von der Raffinesse zugrundeliegender Algorithmen. Die Algorithmen, die heute verwendet werden, existieren in der Theorie teilweise schon seit dem 19. Jahrhundert. Aber erst wenn sehr viele computerlesbare Daten vorhanden sind, kann ein KI-System kalibriert werden und Ergebnisse liefern. Tatsächlich haben die aktuellen Fortschritte im Bereich der KI am wenigsten mit besonders elaborierten Algorithmen zu tun, sondern sind direkt auf die Digitalisierung und die mit ihr einhergehende zunehmende Datafizierung zurückzuführen. Für Bereiche, in denen die Beschaffung computerlesbarer Daten besonders schwierig ist, natürlich umgekehrt, dass der Fortschritt durch KI nur langsam vonstattengehen kann. Dies gilt unter anderem auch in den Rechtswissenschaften, deren Daten hauptsächlich aus geschriebenem Text bestehen - eine, wie noch erläutert werden wird, für Algorithmen besonders schwer zugängliche Sorte Daten.

3

## 2. Keine Künstliche Intelligenz ohne gelabelte Daten

Damit ein Computer Daten lesen kann, müssen diese von einer ganz bestimmten Beschaffenheit sein. Erstens müssen sie in strukturierter Form vorliegen, was in der Informatik einfach ausgedrückt bedeutet, dass sie tabellarisch, mit klar definierten Spalten organisiert sein müssen. Zweitens muss dem Computer sozusagen bekannt sein, welche Bedeutung die Daten, mit denen er gefüttert wird, jeweils repräsentieren. Man spricht in diesem Zusammenhang von sogenannten „gelabelten Daten“. Dieses Labeling erfolgt in Handarbeit und ist daher aufwendig, aber relativ einfach, wenn es zum Beispiel darum geht, einen Algorithmus für die Bilderkennung einzusetzen. Nehmen wir an, ein Algorithmus soll lernen, Bilder zu erkennen, auf denen Katzen zu sehen sind. Hierzu würde der Algorithmus durch eine Vielzahl von Bildern kalibriert, die zuvor mit dem Label „Katze“ versehen wurden. Wird der Algorithmus durch eine große Anzahl derart gelabelter Katzenbilder „trainiert“, wird er schließlich auch auf nicht gelabelten Katzenbildern Katzen erkennen. Alles, was er nicht gelernt hat, etwa Hunde oder Pferde, wird er aber weiterhin nicht bestimmen können.

4

Da nahezu jeder Mensch in der Lage ist, einem Bild das Label „Katze“ zu geben, ist dieses Labeling unkompliziert. Schwieriger wird es, wenn ein Labeling nur von Experten vorgenommen werden kann, beispielsweise bei der radiologischen Befundung von Röntgenbildern, oder eben bei der Kategorisierung juristischer Schriftsätze. Solche hochqualitativ gelabelten Daten haben deshalb einen enormen Wert und wer sie besitzt, hält eine sehr viel wertvollere Ressource als beispielsweise Technologieanbieter, die angeblich intelligente Systeme verkaufen wollen, ohne die hierfür notwendigen Datensätze zu besitzen.

5

Zusammenfassend kann man an dieser Stelle also sagen: Ein Algorithmus ist erstens immer nur so „intelligent“, wie es die Daten, mit denen er kalibriert wurde, zulassen. Die Beschaffung dieser Daten kann je nach Anforderungen einfach oder sehr aufwendig sein. Zweitens geht das, was heute als Künstliche Intelligenz bezeichnet wird, über das Finden von Mustern in Daten und der Übersetzung dieser in Systemverhalten noch nicht hinaus. Das heißt allerdings wiederum nicht, dass KI nicht trotz dieser Beschränkungen ein sehr mächtiges Werkzeug sein kann. Denn erstens arbeitet ein künstlich intelligentes System sehr viel schneller als der Mensch und zweitens findet er Muster auch in einer Detailtiefe und mit einer Genauigkeit, zu der der Mensch nie in der Lage wäre.

6

### **3. In den Rechtswissenschaften tun sich künstlich intelligente Systeme im Moment noch schwer. Warum ist das so?**

Der Annahme, die relativ zögerliche Entwicklung der KI in den Rechtswissenschaften sei hauptsächlich in dem ihr inhärenten Konservatismus begründet, muss an dieser Stelle widersprochen werden. Es ist vielmehr so, dass KI in den Rechtswissenschaften sich mit der für Algorithmen am schwersten zugänglichen Sorte an Daten auseinandersetzen muss: Sprache, bzw. Text. Zwar ist die stark formalisierte Sprache der Rechtswissenschaften oberflächlich betrachtet extrem strukturiert und tatsächlich vereinfacht dieser Umstand die Aufbereitung des Datenmaterials ein wenig. Trotzdem muss betont werden, dass selbst der strukturierteste Schriftsatz für einen Algorithmus ein Wust aus unstrukturierten Daten bleibt, wenn es nicht gelingt, die im Text kodierte Bedeutung in eine maschinell verarbeitbare Repräsentation zu überführen. Diese enorme Herausforderung ist heute aber noch weitestgehend ungelöst.

7

### **4. Natural Language Processing**

Alle vorstellbaren Anwendungen für KI in den Rechtswissenschaften hängen ganz entscheidend davon ab, ob es gelingt, dieser eben genannten Herausforderung zu begegnen. Das ganze Feld der Textkategorisierung ist so anspruchsvoll, dass sich zu diesem Thema eine eigene KI-Subdisziplin entwickelt hat, die man Natural Language Processing, kurz NLP nennt. Was in der juristischen Praxis durch die Methoden von NLP bereits möglich ist, ist zwar durchaus beachtlich, beschränkt sich aber im Wesentlichen auf die Möglichkeiten der Informationsextraktion, also der Extraktion von Bedeutungspartikeln aus juristischen Schriftsätzen, etwa Verträgen. Von einem tatsächlichen Textverständnis sind wir noch weit entfernt. Im Moment werden die verschiedenen Techniken von NLP dazu verwendet, mächtige Assistenzsysteme zu entwickeln, die sich etwa als Suchtechnologie zur Identifizierung relevanter Textstellen oder Dokumente verdient machen. Aber auch die Extrahierung strukturierter Informationen aus Schriftsätzen ist bereits möglich. Nehmen wir beispielsweise Mietverträge. Mit der passenden Software können wir heute quasi auf Knopfdruck aus einer Masse an Mietverträgen eine Tabelle mit den entscheidenden Inhalten - wer mietet

8

was, von wem, ab wann, und wie lange? - erstellen. Ein weiteres großes Feld ist schließlich die Entscheidungsvorhersage und Risikobewertung. Diese Tools befinden sich in einem sehr frühen Entwicklungsstadium, aber es gibt bereits erste Ansätze zur Vorhersage von Gerichtsentscheidungen. Hierfür wird für einen offenen Fall aus einer Datenbank eine möglichst große Menge vergleichbarer, historischer Fälle identifiziert und anhand der relativen Häufigkeit der Erfolgsfälle eine Bewertung abgegeben. Die Entscheidungsvorhersage hat bis jetzt jedoch keine bahnbrechenden Erfolge vorzuweisen und es ist auch fraglich, ob sie abseits von spezifischen Fallkategorien jemals möglich sein wird. Dies könnte allein schon daran scheitern, dass sich sowohl die Gesetzgebung, als auch die richterliche Auslegung im ständigen Fluss befindet und häufig außerdem regionalen Eigenheiten unterliegt.

## 5. Legal Tech - Wohin geht die Reise?

In den nächsten 10 Jahren ist nicht damit zu rechnen, dass Computer juristische Texte verstehen können werden. Es ist allerdings mittelfristig zu erwarten, dass es gelingen wird, einen immer größeren Anteil der Bedeutung von Texten in eine Repräsentation zu überführen, auf dieser Basis algorithmisch zu operieren und so die Vertragsanalyse zur Marktreife zu bringen. Sobald dies gelungen ist, entsteht aus dieser Repräsentation eine Art abstrakte Sprache, ähnlich einer Programmiersprache, mit der der Inhalt von Verträgen in Teilen kodiert werden kann. Die Rechtswissenschaften wären dann endgültig in der Welt der Algorithmen angekommen und dem weiteren Einsatz für Künstliche Intelligenz wären immer weniger Grenzen gesetzt.

9