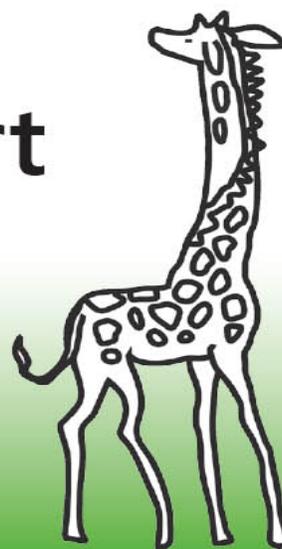
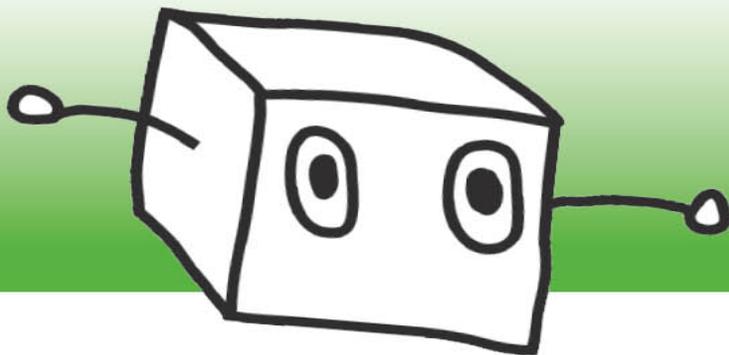


O'REILLY®

Künstliche Intelligenz

Wie sie funktioniert
und wann sie scheitert



Eine unterhaltsame Reise in die seltsame
Welt der Algorithmen, neuronalen
Netze und versteckten Giraffen

Janelle Shane

Übersetzung von Jørgen W. Lang

Papier
plus⁺
PDF.

Zu diesem Buch – sowie zu vielen weiteren O’Reilly-Büchern – können Sie auch das entsprechende E-Book im PDF-Format herunterladen. Werden Sie dazu einfach Mitglied bei oreilly.plus⁺:

www.oreilly.plus

Künstliche Intelligenz

Wie sie funktioniert und wann sie scheitert

Janelle Shane

*Deutsche Übersetzung von
Jørgen W. Lang*

O'REILLY®

Janelle Shane

Lektorat: Alexandra Follenius
Übersetzung: Jørgen W. Lang
Korrektorat: Sibylle Feldmann, www.richtiger-text.de
Satz: III-satz, www.drei-satz.de
Herstellung: Stefanie Weidner
Umschlaggestaltung: Michael Oréal, www.oreal.de

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im
Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:

Print 978-3-96009-160-8
PDF 978-3-96010-495-7
ePub 978-3-96010-496-4
mobi 978-3-96010-497-1

1. Auflage

Translation Copyright für die deutschsprachige Ausgabe © 2021 dpunkt.verlag
GmbH
Wieblinger Weg 17
69123 Heidelberg

Authorized German translation of the English edition of *You Look Like a Thing
and I Love You*, ISBN 9781472268990 © 2019 Janelle Shane. This edition
published by arrangement with Little, Brown and Company, New York, New
York, USA. All rights reserved.

Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem
Imprint »O'REILLY«. O'REILLY ist ein Markenzeichen und eine eingetragene
Marke von O'Reilly Media, Inc. und wird mit Einwilligung des Eigentümers
verwendet.

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger
Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die
Einschweißfolie.



Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen:

komentar@oreilly.de.

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Für alle Leser meines Blogs, die über all die Albernheiten gelacht haben, die die seltsamen Kreaturen gezeichnet und alle Giraffen gefunden haben und die die vom neuronalen Netz erschaffenen Kekse tatsächlich gebacken haben. Danke, dass ihr die Meerrettich-Brownies ausgehalten habt.

Für meine Familie dafür, dass ihr meine größten Fans seid.

Inhalt

Einleitung: KI ist überall

1 Was ist KI?

Klopf klopf, wer ist da?
Die KI einfach mal machen lassen
Manchmal sind die Regeln schuld
Wie man eine schlechte Regel erkennt
Vier Anzeichen für KI-Katastrophen
Fluch oder Segen?

2 KI ist überall, aber wo genau ist das? - Warum KI für bestimmte Aufgaben besser geeignet ist

Dieses Beispiel ist wirklich wahr, ehrlich!
Für mich wäre es eigentlich völlig in Ordnung, wenn ein Roboter das für mich macht
Je enger eine Aufgabe umrissen ist, desto schlauer ist die KI
C-3PO und Ihr Toaster: ein Intelligenzvergleich
Unzureichende Daten führen zu fehlerhaften Berechnungen
Gelerntes wiederverwenden
Erinnerung? Vergessen Sie's!
Kann man das Problem auch einfacher lösen?
Darf eine KI Auto fahren?

3 Wie lernt eine Maschine eigentlich wirklich? - KI-Arten, ihre Funktionsweisen und Eigenarten

Neuronale Netzwerke
Das magische Sandwich-Portal
Der Trainingsprozess
Wenn Zellen zusammenarbeiten
Markow-Ketten
Random Forest - Zufallswälder
Evolutionäre Algorithmen

Generative gegnerische Netze (GANs)
Mischen, abstimmen und zusammenarbeiten

4 Ich versuch's doch! - Warum alles vom Datensatz abhängt

Zu weit gefasste Probleme
Bitte mehr Daten
Ungenauere Daten
Daten, die Zeit verschwenden
Is this the real life?
Andere Eigentümlichkeiten von Datensätzen
Fehlende Daten
Ich sehe vier Giraffen

5 Worum geht es wirklich? - Die KI löst das falsche Problem

Belohnungsfunktionen hacken
Computerspiele sind verwirrend
Bitte nicht gehen!
Neugier
Hüten Sie sich vor fehlerhaften Belohnungsfunktionen

6 Die Matrix hacken - Wenn die KI Fehler in der Simulation ausnutzt

Du hast nicht gesagt, dass ich das nicht darf
Mathematische Fehler zum Abendessen
Viel mächtiger, als Sie sich es jemals vorstellen können

7 Unglückliche Abkürzungen - Verzerrte Ergebnisse durch Überanpassung und Vorurteile

Klassenungleichheit
Überanpassung
Das Hacken der Matrix funktioniert nur in der Matrix
Menschen imitieren
Keine Empfehlung, sondern eine Vorhersage
Die Ergebnisse der KI überprüfen

8 Ist eine KI aufgebaut wie ein menschliches Gehirn? - Ähnlichkeiten und entscheidende Unterschiede

KI-Traumwelten

Echte und künstliche Gehirne denken auf die gleiche Weise

Katastrophale Vergesslichkeit

Verstärkung von Vorurteilen

Gegnerische Angriffe

Das Offensichtliche übersehen

9 Menschliche Bots - Wo Sie KI vermutlich nicht finden werden

Ein als Roboter verkleideter Mensch

Bot oder nicht?

10 Eine Partnerschaft zwischen Mensch und KI

Instant-KI: einfach menschlichen Sachverstand hinzufügen

Wartung

Hüten Sie sich vor KIs, die im laufenden Betrieb lernen

Ein klarer Job für die KI

Algorithmische Kreativität?

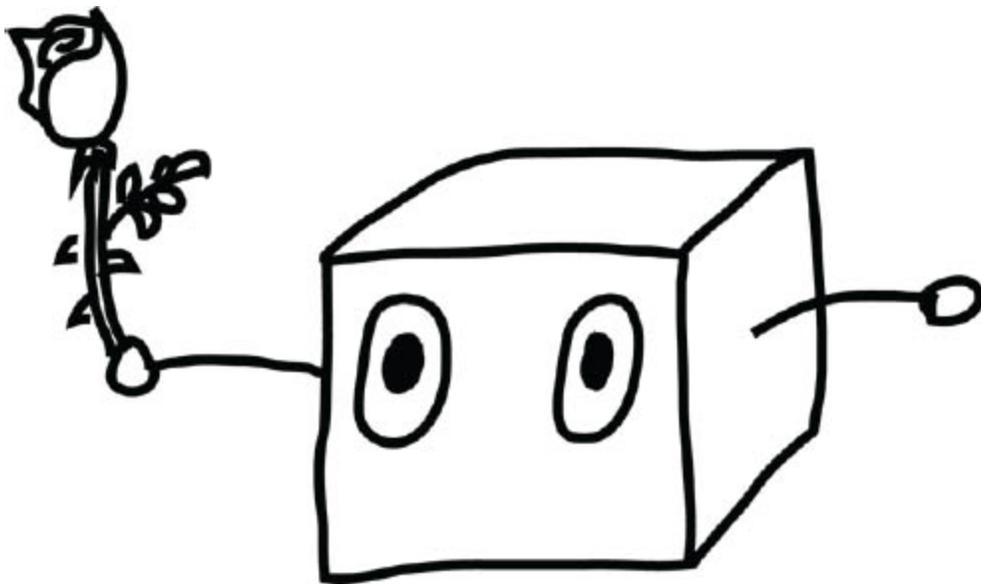
Fazit: Das Leben mit unseren künstlichen Freunden

Danksagungen

Anmerkungen

Index

KI ist überall



Einer künstlichen Intelligenz (KI) beizubringen, wie man flirtet, ist auch für mich keine alltägliche Aufgabe.

Und dabei habe ich schon eine Menge seltsamer KI-Projekte durchgeführt. In meinem Blog *AI Weirdness* (<https://aiweirdness.com/>) habe ich eine KI beispielsweise darauf trainiert, sich Katzennamen auszudenken. Die Namen Mr. Tinkles («Mr. Pinkel») und Retchion («Würgion») waren dabei eher weniger erfolgreiche Versuche. Auch hatte ich schon probiert, einer KI beizubringen, wie man neue Kochrezepte erfindet. Dafür waren dann Zutaten wie geschälter Rosmarin oder eine

Handvoll zerbrochenes Glas nötig. Aber der Versuch, einer KI das Flirten beizubringen, war noch einmal etwas vollkommen anderes.

KI lernt anhand von Beispielen. In meinem Fall hieß das, eine Liste mit Flirtsprüchen zu studieren und darauf basierend neue Sprüche zu erzeugen. Das Problem dabei war, dass ich die Trainingsdaten auf meinem Computer aus verschiedenen ziemlich furchtbaren Quellen im Internet zusammengesucht hatte. Sie reichten von anzüglichen Wortspielen bis hin zu ziemlich derben sexuellen Anspielungen. Sobald ich die KI darauf trainiert haben würde, diese Sprüche zu imitieren, sollte sie auf Knopfdruck Tausende neuer Varianten produzieren können. Wie ein vollkommen unbefangenes Kind würde sie absolut keine Ahnung haben, was sie imitieren sollte und was sie besser vermeidet. Die KI würde komplett bei null anfangen, ohne zu wissen, was Anmachsprüche sind (oder auch nur, was Englisch ist), und aus den Beispielen lernen. Dabei würde sie ihr Bestes geben, um die gefundenen Muster zu imitieren - inklusive der Grobheiten. Woher soll sie es auch besser wissen?

Ich war schon drauf und dran, das Projekt aufzugeben, aber ich musste einen Blogartikel schreiben und hatte schon viel zu viel Zeit mit dem Sammeln der Beispielsprüche verbracht. Also begann ich mit dem Training. Die KI suchte nach Mustern, sie erfand und testete Regeln, die ihr dabei helfen sollten, vorherzusagen, welche Buchstaben in einem Anmachspruch in welcher Reihenfolge vorkommen sollten. Endlich war das Training beendet. Etwas beklommen forderte ich die KI auf, ein paar Zeilen auszugeben:*

Bist du eine Gardinenstange? Denn du bist das Einzige, was es

hier gibt.

Hey Baby, du musst ein Schlüssel sein, denn ich kann dein
Tuten

ertragen.

Bist du eine Kerze? Weil du so heiß bist von den Blicken mit
dir.

Du bist so schön, dass du sagst, eine Fledermaus auf mich
und Baby.

Du siehst aus wie ein Ding und ich liebe dich.

Ich war überrascht - und erleichtert. Das virtuelle Gehirn der KI (ungefähr so komplex wie das eines Wurms) war nicht in der Lage, irgendwelche Feinheiten, einschließlich Frauenfeindlichkeiten oder Geschmacklosigkeiten, in den Trainingsdaten aufzuspüren. Sie hat die gefundenen Muster so gut sie konnte imitiert. Auf diese Weise hat sie eine deutlich bessere Lösung für das Problem gefunden - nämlich einen Fremden zum Lachen zu bringen.

Für mich waren diese Sprüche ein durchschlagender Erfolg. Wenn Ihr Wissen über künstliche Intelligenz dagegen aus Zeitungsüberschriften und Science-Fiction-Romanen stammt, kann diese Ahnungslosigkeit der KI jedoch sehr überraschend sein. Unternehmen behaupten gern, dass KI die Feinheiten menschlicher Sprache mindestens so gut wie Menschen beurteilen kann. Sie vermitteln den Eindruck, der Mensch werde in den meisten Berufen schon sehr bald durch künstliche Intelligenz ersetzt werden. In ihren Pressemitteilungen kann man lesen, die KI würde in allernächster Zukunft alle Lebensbereiche durchdrungen haben. Einerseits haben sie damit recht - andererseits aber absolut nicht.

In Wirklichkeit ist KI schon jetzt überall. Sie beeinflusst, was wir online erleben, legt fest, welche Werbeanzeigen wir zu sehen bekommen, und schlägt uns Videos vor. Daneben spürt sie Social-Media-Bots auf und warnt uns vor böartigen Websites. Unternehmen setzen KI ein, um anhand von Lebensläufen zu entscheiden, welche Kandidaten zum Bewerbungsgespräch eingeladen werden, wer kreditwürdig ist und wer nicht. KI-gesteuerte selbstfahrende Autos haben bereits Millionen von Kilometern zurückgelegt (wobei in verwirrenden Situationen immer noch Menschen eingreifen). Auch in unseren Smartphones ist KI am Werk. Sie reagiert auf Sprachbefehle, markiert automatisch die Gesichter in unseren Fotos und setzt uns in Videos diese süßen Häschenohren auf.

Aus Erfahrung wissen wir aber auch, dass künstliche Intelligenz nicht ohne Fehler ist – bei Weitem nicht. Wer kennt das nicht: Ständig wird uns Werbung für Schuhe angezeigt, die wir längst gekauft haben. Spamfilter übersehen Betrugsversuche, während die wichtigsten E-Mails zur ungünstigsten Zeit im Spamordner landen.



Je mehr unser Leben von Algorithmen bestimmt wird, desto größer und weitreichender sind die Konsequenzen, wenn die KI fehlerhaft ist. Die Empfehlungsalgorithmen von YouTube leiten uns zu immer stärker polarisierenden Inhalten. Schon mit wenigen Klicks gelangen wir von glaubwürdigen Nachrichtenquellen zu Videos, die Hassreden und Verschwörungserzählungen enthalten.¹ Algorithmen, die entscheiden, wer auf Bewährung freikommt, einen Kredit erhält oder zum Bewerbungsgespräch eingeladen wird, sind nicht unparteiisch. Sie unterliegen den gleichen Vorurteilen wie die Menschen, die sie eigentlich ersetzen sollen – manchmal ist diese Voreingenommenheit sogar noch deutlich stärker. Eine KI-basierte Überwachungsfunktion ist unbestechlich, sie kann aber auch keine moralischen Bedenken zu den übertragenen Aufgaben äußern. Auch kann sie Fehler machen, wenn sie falsch benutzt oder gar gehackt wird. Forscher haben entdeckt, dass schon ein unscheinbarer kleiner Aufkleber eine Bilderkennungs-KI dazu bringen kann, eine Pistole für einen Toaster zu halten; ein Fingerabdruckscanner mit niedrigem Sicherheitsstandard konnte in 77 Prozent aller Fälle mit einem einzigen Master-Fingerabdruck ausgetrickst werden.

Oft wird KI als deutlich leistungsfähiger angepriesen, als sie es tatsächlich ist. Manche Unternehmen behaupten Dinge, die klar ins Reich der Science-Fiction gehören. Andere bewerben ihre KI als unbestechlich, obwohl ihr Verhalten messbar diskriminierend ist. Oft ist das, was als Leistung einer KI angepriesen wird, in Wirklichkeit die Arbeit von Menschen, die hinter den Kulissen tätig sind. Als Konsumenten und Bürger dieses Planeten dürfen wir uns nicht übertölpeln lassen. Wir müssen verstehen, wie unsere Daten verwendet werden und was die von uns benutzte künstliche Intelligenz tatsächlich ist – und was nicht.

In meinem Blog *AI Weirdness* führe ich aus Spaß Experimente mit künstlicher Intelligenz durch. Manchmal lasse ich KI Dinge imitieren, die eher unüblich sind, zum Beispiel Anmachsprüche. In anderen Fällen versuche ich, die KI an ihre Grenzen zu bringen. Einmal habe ich einer KI beispielsweise ein Bild von Darth Vader gezeigt und gefragt, was sie sieht. Sie behauptete, Darth Vader sei ein Baum, und versuchte dann, mit mir darüber zu diskutieren. Durch meine Experimente fand ich heraus, dass eine KI selbst an den einfachsten Aufgaben scheitern kann, als hätte man sie absichtlich auf die falsche Fährte gelockt. Gerade dadurch, dass man einer KI eine Aufgabe gibt und beobachtet, wie sie versagt, kann man eine Menge über sie lernen.

Wie wir in diesem Buch sehen werden, sind die Dinge, die im Innern von KI-Algorithmen vor sich gehen, oft sehr seltsam und irgendwie ineinander verschlungen. Ihr Output ist häufig die einzige Quelle, mit deren Hilfe man herausfinden kann, was sie verstanden hat und was vollkommen schiefgegangen ist. Weisen Sie eine KI an, eine Katze zu zeichnen oder sich einen Witz auszudenken, macht sie die gleiche Art von Fehlern wie bei der Analyse von Fingerabdrücken oder der Sortierung medizinischer Abbildungen. Der Unterschied ist, dass man leichter erkennen kann, dass etwas nicht stimmt, weil die Katze sechs Beine hat oder dem Witz die Pointe fehlt. Außerdem machen diese Experimente einen Heidenspaß.

Im Verlauf meiner Versuche, KIs aus ihrer Komfortzone herauszuholen und unserer näherzubringen, habe ich KIs angewiesen, die erste Zeile eines Romans zu schreiben, Schafe an ungewöhnlichen Orten zu erkennen, Rezepte zu schreiben, Namen für Meerschweinchen zu erfinden und sich überhaupt so seltsam wie möglich zu verhalten. Durch diese Experimente können Sie lernen, was KI wirklich gut

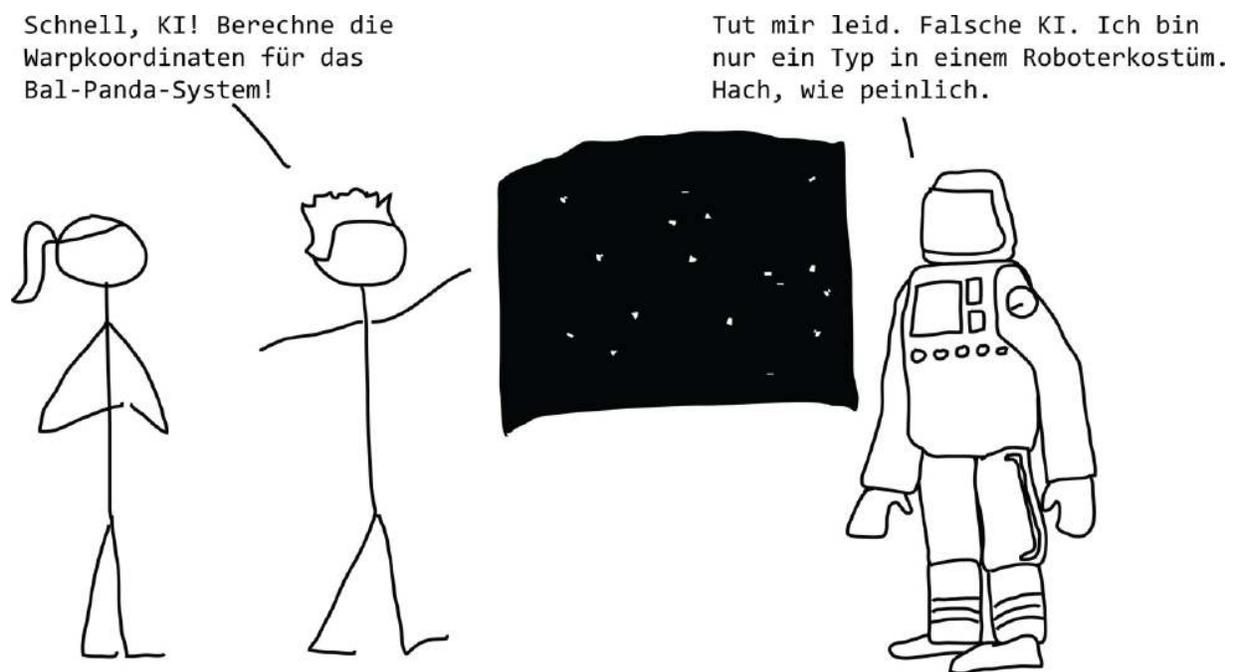
kann und wo sie Schwierigkeiten hat – und was sie weder in meinem noch in Ihrem Leben vermutlich nie lernen wird.

Die folgenden fünf Prinzipien der KI-Seltsamkeit habe ich herausgefunden:

- Die Gefahr ist nicht, dass eine KI zu schlau ist, sondern dass sie nicht schlau genug ist.
- Die »Intelligenz« einer KI entspricht ungefähr der eines Wurms.
- KI versteht die zu lösenden Probleme nicht wirklich.
- Aber: Eine KI macht genau das, was Sie ihr sagen. Zumindest versucht sie, ihr Bestes zu geben.
- Die KI geht den Weg des geringsten Widerstands.

Und damit wollen wir eintauchen in die seltsame Welt der künstlichen Intelligenz. Wir werden lernen, was KI ist – und was nicht. Wir werden lernen, worin sie gut ist und wo sie zum Scheitern verurteilt ist. Wir werden lernen, warum zukünftige KIs mehr Ähnlichkeit mit einem Insektenschwarm haben als mit C-3PO. Wir werden lernen, warum ein selbstfahrendes Auto bei einer Zombie-Apokalypse ein ziemlich schlechtes Fluchtfahrzeug ist. Außerdem werde ich Ihnen zeigen, warum Sie sich niemals freiwillig melden sollten, um eine KI zu testen, die Sandwiches sortiert, und Sie werden laufende KIs kennenlernen, die alles gerne tun außer zu laufen. Auf diese Weise lernen wir die Funktionsweise von KIs kennen: wie sie denkt und warum die Welt dadurch zu einem seltsameren Ort wird.

Was ist KI?



Man könnte glauben, dass KI bereits alle Lebensbereiche durchdrungen hat. Das liegt zum Teil daran, dass der Begriff »künstliche Intelligenz« so viele verschiedene Bedeutungen hat. Je nachdem, ob Sie Science-Fiction-Literatur lesen, eine neue App verkaufen oder wissenschaftliche Forschung betreiben, wird er anders interpretiert. Wenn jemand sagt, er habe einen Chatbot mit KI-Fähigkeiten, können wir dann erwarten, dass er Meinungen und Gefühle hat wie C-3PO? Oder ist es nur ein

Algorithmus, der gelernt hat, zu erraten, wie Menschen wahrscheinlich auf einen bestimmten Satz reagieren? Oder eine Tabelle, die Wörter in Ihrer Frage mit vorformulierten Antworten abgleicht? Oder ein unterbezahlter Mensch, der alle Antworten von irgendeinem Ort der Welt von Hand eintippt? Oder vielleicht sogar eine komplett vorgefertigte Unterhaltung, in der Mensch und KI vordefinierte Textzeilen vorlesen wie die Figuren in einem Theaterstück? Verwirrenderweise wurden alle diese Definitionen schon benutzt, um zu erklären, was KI bedeutet.

In diesem Buch benutzen wir den Begriff *künstliche Intelligenz (KI)* so, wie er heutzutage üblicherweise von Programmierern verwendet wird: als eine bestimmte Art von Computerprogramm, die als **Algorithmus für maschinelles Lernen** bezeichnet wird. Das folgende Diagramm zeigt eine Reihe von Begriffen, die wir in diesem Buch behandeln und wie sie nach dieser Definition einzuordnen sind.

Als KI bezeichnete Dinge

KI in diesem Buch:

Algorithmen für maschinelles Lernen
 Deep Learning
 Neuronale Netzwerke
 Rekurrente neuronale Netzwerke
 Markow-Ketten
 Random Forests
 Genetische Algorithmen
 Generative Adversarial Networks
 Reinforcement Learning
 Smartphone-Textvorschläge
 Magische Sandwich-Sortierer
 Unglückliche Mörder-Roboter

In diesem Buch, aber keine KI:

Science-Fiction-KIs
 Regelbasierte Programmierung
 Als Roboter verkleidete Menschen
 Roboter, die Skripte vorlesen
 Menschen, die vorgeben, KIs zu sein
 Kakerlaken mit Bewusstsein
 Phantomgiraffen



Alles, was in diesem Buch als KI bezeichnet wird, ist auch ein Algorithmus für maschinelles Lernen. Daher wollen wir zunächst einmal sehen, was das eigentlich ist.

Klopf klopf, wer ist da?

Um eine KI in freier Wildbahn zu erkennen, müssen Sie zuerst den Unterschied zwischen **Algorithmen für maschinelles Lernen** (die wir in diesem Buch als KI bezeichnen) und traditionellen Programmen kennen (die Programmierer üblicherweise als **regelbasiert** bezeichnen). Wenn Sie sich jemals mit einfacher Programmierung beschäftigt oder auch nur eine HTML-Seite erstellt haben, werden Sie mit großer Sicherheit ein regelbasiertes Programm benutzt haben. Sie erstellen eine Liste mit Befehlen oder Regeln in einer Computersprache, und der Computer tut genau das, was Sie ihm gesagt haben. Um mit einem regelbasierten Programm ein Problem zu lösen, müssen Sie alle dafür nötigen Schritte kennen und beschreiben können.

Ein Algorithmus für maschinelles Lernen findet die Regeln dagegen durch selbstständiges Ausprobieren, indem er seine Fortschritte ständig mit dem vom Programmierer vorgegebenen Ziel abgleicht. Das kann eine Liste mit Beispielen sein, die imitiert werden sollen, ein Spielstand, der verbessert werden soll, oder etwas vollkommen anderes. Während die KI versucht, ihr Ziel zu erreichen, kann sie sogar Regeln und Beziehungen finden, deren Existenz der Programmierer nicht einmal geahnt hat. Die Programmierung einer KI hat mehr Ähnlichkeit damit, einem Kind etwas beizubringen, als damit, einen Computer zu programmieren.

Regelbasierte Programmierung

Angenommen, wir wollten regelbasierte Programmierung benutzen, um einem Computer beizubringen, wie man Klopf-klopf-Witze* erzählt. Zuerst müssen wir herausfinden, nach welchen Regeln die Witze funktionieren. Ich würde ihre Struktur analysieren und feststellen, dass sie nach einer Art Formel ablaufen:

Klopf klopf.

Wer ist da?

[Name/Begriff]

[Name/Begriff] Wer?

[Name/Begriff] [Pointe]

Sobald wir diese Formel kennen, gibt es nur noch zwei Stellen, die das Programm erzeugen muss: [Name/Begriff] und [Pointe]. Aber auch hierfür werden Regeln gebraucht.

Ich könnte eine Liste mit Namen/Begriffen und dazu passenden Pointen erstellen, wie etwa diese:

Namen/Begriffe Pointen

| | |
|-------------------|--|
| <i>Mani</i> | <i>Manitu. Für Mani tu ich alles.</i> |
| <i>Muh</i> | <i>Muh-Tation.</i> |
| <i>Salat</i> | <i>Klopf-Salat.</i> |
| <i>Werner</i> | <i>Wer nervt mich andauernd mit diesen Klopf-klopf-Witzen?</i> |
| <i>Karla</i> | <i>Karla Gerfeld.</i> |
| <i>Tupper</i> | <i>Tupperware.</i> |
| <i>Australien</i> | <i>Kängurus kommen Aus Tralien.</i> |

Jetzt kann der Computer einen Klopf-klopf-Witz erzeugen, indem er ein Paar aus Name und Pointe auswählt und in die Vorlage einfügt. Hierbei werden allerdings *keine neuen Witze* erzählt, sondern nur solche, *die wir bereits kennen*. Ich könnte versuchen, das etwas interessanter zu machen, indem wir [Manitu] zum Beispiel durch [Manipulation] ersetzen. Dann kann das Programm einen neuen Witz erzeugen:

Klopf klopf.

Wer ist da?

Mani

Mani wer?

Manipulation. Du wirst jetzt ganz müde.

Ich könnte [Manitu] durch [Manifest], [Manila] oder irgendetwas anderes ersetzen, und der Computer könnte noch mehr neue Witze produzieren. Mit genug Regeln könnte ich vermutlich Hunderte neuer Witze ausgeben lassen.

Je nachdem, wie ausgeklügelt das Programm sein soll, könnte ich sehr viel Zeit mit dem Festlegen neuer Regeln verbringen. Ich könnte eine Liste mit vorhandenen Pointen finden und nach einer Möglichkeit suchen, sie in das nötige Format für die Klopf-klopf-Witze umzuwandeln. Ich könnte sogar versuchen, Ausspracheregeln zu programmieren oder Reime, Homophone oder kulturelle Anspielungen und

so weiter. Diese könnte der Computer dann zu neuen interessanten Pointen zusammensetzen. Wenn ich das schlaue genug anstelle, kann das Programm Pointen erstellen, die es nie zuvor gehört hat. (Allerdings hat eine Person, die das tatsächlich ausprobiert hat, festgestellt, dass die vom Algorithmus verwendeten Wörter und Phrasen so altmodisch und verworren waren, dass sie heutzutage niemand mehr verstehen würde.) Dennoch, egal wie ausgeklügelt meine Witzeregeln auch sein mögen: Ich gebe dem Computer immer genau Anweisungen dazu, wie er das Problem lösen soll.

KI trainieren

Wenn ich dagegen eine KI darauf trainiere, Klopf-klopf-Witze zu erzählen, lege ich keine Regeln fest. Die KI muss die Regeln selbst herausfinden.

Ich übergebe ihr nur eine Reihe schon bekannter Klopf-klopf-Witze und ein paar absolut notwendige Spielregeln: »Hier hast du ein paar Klopfklopf-Witze. Versuche, mehr davon herzustellen.« Und das Rohmaterial, mit dem die KI arbeiten soll? Ein Sack voll zufälliger Buchstaben und Satzzeichen.

Danach hole ich mir erst mal einen Kaffee.

Die KI macht sich an die Arbeit.

Zuerst versucht sie, ein paar Buchstaben zu finden, die in Klopfklopf-Witzen vorkommen. Im Moment sind die Rateversuche dabei komplett zufällig. Der erste Versuch könnte also alles Mögliche sein, zum Beispiel: »qasdnw,m sne?mso d.« Soweit die KI im Moment weiß, wird so ein Klopf-klopf-Witz erzählt.

Dann sieht sich die KI an, was Klopf-klopf-Witze *tatsächlich* sein sollen. Sehr wahrscheinlich liegt sie dabei ziemlich falsch. »Also gut«, sagt sich die KI und passt ihre eigene Struktur ein wenig an, damit sie beim nächsten Mal

etwas besser rät. Dabei ist die Stärke der Änderungen begrenzt, um zu verhindern, dass sich die KI jedes neu gefundene Textstück merkt. Aber schon mit minimalen Anpassungen findet die KI heraus, dass sie zumindest nicht vollkommen falsch liegt, wenn sie nur *ks* und Leerzeichen errät. Nachdem sie sich den ersten Stapel Klopf-klopf-Witze angesehen und ein paar Korrekturen vorgenommen hat, ist das nun Folgende ihre Vorstellung davon, wie ein Klopf-klopf-Witz auszusehen hat:*

k k k k k
kk k kkkok
k kkkk
kk
kk k kk
klokp k
k
k

Das ist jetzt nicht gerade der beste Klopf-klopf-Witz der Welt. Aber mit diesem Startpunkt kann die KI mit einem zweiten Stapel Witze weitermachen und dann mit noch einem. Bei jedem Durchgang verfeinert sie ihre Witzeformel weiter, um zu immer besseren Ergebnissen zu gelangen.

Nach ein paar Runden mit Rateversuchen und Anpassungen mehr hat die KI weitere Regeln gelernt. Sie

weiß nun, dass am Ende einer Zeile gelegentlich ein Fragezeichen auftauchen muss. Sie lernt, Vokale zu benutzen (insbesondere *o*). Sie versucht sich sogar an der Verwendung von Kommata.

noo,

Lnop noo

Kof?

hnos h st

wrst oa , a

asutWen

klo ada

pf kla

w is

e

Wie gut stimmen ihre Regeln für die Erzeugung von Klopf-klopf-Witzen mit der Realität überein? Offenbar fehlt noch etwas.

Um einen guten Klopf-klopf-Witz zu erzählen, muss sie herausfinden, in welcher *Reihenfolge* die Buchstaben angeordnet werden müssen. Auch hier beginnt sie mit ein paar Rateversuchen. Folgt auf ein *o* immer ein *w*? Offenbar keine so gute Idee. Aber dann rät sie, dass auf ein *o* oft *pf* folgt. Großartig. Wieder hat die KI einen Fortschritt erzielt. So sieht für sie aktuell der perfekte Witz aus:

Weopf

Weopf

Weopf

Weopf

Weopf Weopf Weopf

Weopf Weopf

Weopf

Weopf

Das ist immer noch kein Klopf-klopf-Witz – es klingt mehr nach einem Huhn mit einem Sprachfehler. Die KI muss also weitere Regeln finden.

Sie sieht sich die Trainingsdaten erneut an und sucht nach neuen Möglichkeiten, »opf« zu benutzen. Dabei versucht sie, Kombinationen zu finden, die noch besser auf die bestehenden Beispiele für Klopf-klopf-Witze passen.

klpf wopf worpf

Wkl wWlopf

Kmopf

er is Westa Wler

looo ooop

Keee?

eerr

lop lo,p Wler s rsit
ea lo oo pf KropfWnopf Woors
Dapf
lop K opf
Kop
wee
KKopf Klopf Daopf Weompf

Diese Entwicklung dauert nur wenige Minuten. Als ich mit meinem Kaffee zurückkomme, hat die KI schon herausbekommen, dass es sehr, sehr gut zu den Beispielen passt, wenn die Witze mit »Klopf klopf, wer ist da?« beginnen. Sie entscheidet sich, diese Regel *immer* anzuwenden.

Allerdings dauert es etwas länger, den Rest der Formel herauszubekommen. Dabei verfängt sich die KI immer wieder in einer Art Schleife, in der sie mit sich selbst »Stille Post« spielt.

Klopf klopf.
Wer ist da?
lien
Austra wer?
Aaani wer?
Alien wer?
Aulali wer?

trali wer?

stilien wer?

Schon bald bekommt sie auch den Rest der Formel heraus, allerdings noch ohne richtige Wörter (die KI hat keine Ahnung, was »Wörter« eigentlich sind):

Klopf klopf.

Wer ist da?

Wern

Hane wer?

Glarm sind dann denen da

Manchmal ist die Pointe auch etwas zu lang:

Klopf klopf.

Wer ist da?

Dieelieb

Manipizz wer?

Hingeh den letzt emotiwan aut ehn weonges eich anet ehn tee

soor dein onithit das wo dein thasg glib tine wenn dasle
ter

den dieung!

Irgendwann im Verlauf des Trainings lernt die KI den Witz mit der lippenlosen Kuh, der ab sofort ihr »Absolut Großartigster Lieblingwitz Aller Zeiten« ist. Falls Sie den Witz nicht kennen – hier ist er:

Klopf klopf.

Wer ist da?

Eine lippenlose Kuh.

Welche lippenlose Kuh?

Eine lippenlose Kuh macht uuh, uuuuuh, uuuuh.

Für die KI ist dies aktuell die höchste Form des Humors. Also glaubt die KI zunächst für einige Zeit, dass die Pointe jedes Klopf-klopf-Witzes mit *uuh, uuuuuh, uuuuh* endet.

Klopf klopf.

Wer ist da?

Alex

Alim wer?

Wuh ku zuh den do do duhhr zu dem suuuh uuu uuuu uuh uu uu
uu

uuuuuu uuh uuuuuh uuu uuu uuh uuuu uuu uu...

Okay, die KI hat eine neue Regel gefunden. Nach und nach (und vermutlich etwas widerstrebend) nimmt sie wieder etwas Abstand von der »Kuh ohne Lippen«-Pointe. Sie hat aber weiterhin Probleme, die eine erstaunliche Ähnlichkeit mit zwei Personen haben, die sich Klopf-klopf-Witze über

eine schlechte Telefonverbindung erzählen und dabei offenbar vergessen, wer denn jetzt den Witz erzählt:

Klopf klopf.

Wer ist da?

Alter

Wasser wer?

Wasser wer wer?

Wer wer?

Wasser damit macht gerne blau?

Schließlich findet die KI die korrekte Struktur eines Klopfklopf-Witzes. Jetzt erzeugt sie Witze, die technisch zwar funktionieren, aber immer noch wenig Sinn ergeben. Viele der Witze scheinen Plagiate aus den Trainingsdaten zu enthalten:

Klopf klopf.

Wer ist da?

Mani

Mani wer?

Raubüberfall! Her mit der Knete!

Klopf klopf.

Wer ist da?

Salat

Salat wer?

Eine lippenlose Kuh kann gar nicht sprechen.

Klopf klopf.

Wer ist da?

Australien

Australien wer?

Aus Tralien Gerfeld.

Und dann ... produzierte die KI ... tatsächlich einen Witz ... selbstständig, ohne hierfür auf die Trainingsdaten zurückzugreifen. Der Witz war nicht nur verständlich, sondern sogar ... lustig?

Klopf klopf.

Wer ist da?

Mark

Mark wer?

Ich Mark Klopf-klopf-Witze.

Hat die KI also tatsächlich begonnen, Klopf-klopf-Witze und Wortspiele zu verstehen? Bei dem relativ kleinen Trainingsdatensatz ist das mehr als unwahrscheinlich.