

## Machine Learning – Nutzen und Tücken von Algorithmen in der Kommunikation

Künstliche Intelligenz via Machine Learning ist einer der zentralen Punkte einer Customer Data Platform (CDP). Funktionen wie die erweiterte Kundenanalyse, eine kontextbezogene Personalisierung oder die Kundenkommunikation in Echtzeit werden dadurch erst möglich.

In der Customer Data Platform laufen alle Kundendaten zentral zusammen. Doch das allein ist nur der erste Schritt. Erst wenn die Daten ausgelesen und interpretiert werden können, fängt der wahre Nutzen an. Der Zugriff auf leistungsstarke datenwissenschaftliche maschinelle Lernalgorithmen erleichtert Marketern das Leben erheblich. Im Machine Learning lesen selbstlernende Systeme auf Basis von Algorithmen aus dem Verhalten der Kunden. Auch externe Datenquellen, wie beispielsweise das Wetter oder demografische Daten aus der Region oder dem Wohnumfeld der Kunden, können miteinbezogen werden.

Das System lernt ständig, ohne dass im Marketing willkürlich Regeln vordefiniert oder und definiert werden müssen. Denn wir Menschen können einfach nicht damit konkurrieren, wie schnell Maschinen lernen. Und so werden Kundenverhaltenstrends und Vorhersageinformationen sehr schnell und einfach möglich.

Jedoch lauern beim maschinellen Lernen auch echte Gefahren, da die Algorithmen mit den Daten arbeiten, die sie bekommen. Wenn diese Daten nicht ausgewogen definiert sind oder Definitionslücken enthalten, kann es zu bösen Überraschungen, unerwünschten Ergebnissen und Filterblasen kommen.

### Algorithmen im Alltag

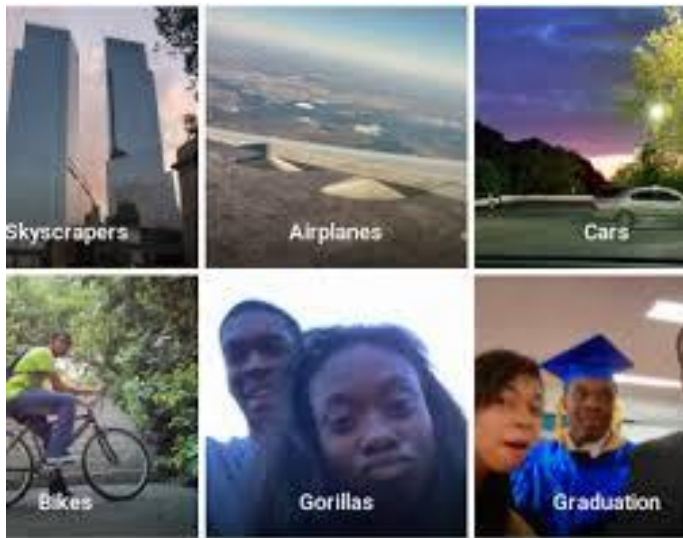
Algorithmen beeinflussen unser Leben im digitalen Zeitalter auf unterschiedlichste Weise. Amazon schlägt basierend auf zurückliegenden Käufen passende Artikel vor, Google vervollständigt Suchanfragen und Facebook platziert Werbung, die den Nutzer am wahrscheinlichsten interessiert.

Im Marketing können Scorings die Entscheidung treffen, für welche Kunden das knappe Budget ausgegeben werden sollte, und welche Neukunden sich am wahrscheinlichsten zu den neuen Stammkunden entwickeln. Algorithmen können Hochrechnungen für Produkte oder die Platzierung von Online-Werbung optimieren oder eine neutrale Bewertung zu Bewerbern abgeben.

Doch für Betroffene können Nachteile entstehen. Ein schlechtes Rating der Schufa kann dazu führen, dass ein Kredit nicht vergeben wird. Dabei sind viele Algorithmen für den Betroffenen, und häufig auch für die Entwickler, nicht zu durchschauen.

## Wie Algorithmen diskriminieren

Bereits 2015 ging durch die Medien: „Google-Fotos bezeichnet Schwarze als Gorillas“. Bilder eines dunkelhäutigen Paares waren automatisch in einen Ordner mit dem Namen „Gorilla“ verschoben worden. Zuvor war Flickr etwas Ähnliches passiert. Der Algorithmus zum automatischen Tagging von Fotos hatte vermehrt Fotos von dunkelhäutigen Menschen mit Tags wie „Affe“ und „Tier“ versehen. Wie kann es dazu kommen, dass ein Algorithmus sich scheinbar rassistisch verhält?



Quelle: jackyalcine@twitter

Zunächst sind Computer nicht rassistisch. Sie lernen aus den Daten, die sie bekommen und auf eine vorgegebene Weise. Sind nun aber die Trainingsdaten nicht sauber definiert, so kann der Algorithmus daraus auch unbeabsichtigt etwas Ungewünschtes lernen. Enthalten die Daten zum Beispiel nur weiße Gesichter, so kann er nicht lernen, dunkelhäutige Gesichter als solche zu erkennen. Er muss sich für eine andere gelernte Kategorie entscheiden, die ihm am treffendsten erscheint. Dass der Algorithmus dabei auf die gelernte Kategorie „Gorilla“ zurückgreift, ist nicht vorsätzlich.

Das Problem liegt also nicht im Algorithmus, sondern in der Definition der Daten. Den Entwicklern ist entgangen, dass ihre Trainingsdaten offensichtlich zu wenige Bilder von dunkelhäutigen Menschen enthalten haben, und diese somit nicht gelernt werden konnten. Auch hier liegt sehr wahrscheinlich kein Vorsatz vor, sondern die Nachlässigkeit der Programmierer, welchen die Unausgewogenheit nicht aufgefallen ist. Es kann nur vermutet werden, dass ein Algorithmus, der hellhäutige Gesichter systematisch als Tiere erkannt hätte, nicht publiziert worden wäre.

Nicht nur bei Gesichtern hat Bilderkennungs-Software Probleme, auch die Unterscheidung von Babys und Robben oder Hunden und Katzen ist für einen Algorithmus schwierig. Was einem Menschen intuitiv sehr einfach gelingt, muss ein Computer mühsam an einer großen und möglichst vollständigen Menge an Daten trainieren.

Nicht nur Ethnien, auch Geschlechter können durch künstliche Intelligenz diskriminiert werden. Ein Beispiel dafür ist der Recruiting-Algorithmus, den Amazon zum Einsatz bringen wollte. Der Ansatz war, Bewerber neutral zu bewerten um so, ohne großen Aufwand, die vielversprechendsten Bewerbungen herauszufiltern. Personalern hätten so im Bewerbungsprozess mehr Zeit für diese am ehesten passenden Bewerber. So einleuchtend diese Idee auch klingen mag, so vorsichtig muss sie umgesetzt werden.

Unter den Bewerbern bei Amazon, die als Trainingsdatensatz herangezogen wurden, waren nur wenige Frauen zu finden. Daraus musste der Algorithmus schließen, dass Frauen keine geeigneten Angestellten sein können. Selbst wenn das Geschlecht nicht explizit angegeben wird oder bewusst aus der Auswertung als Merkmal ausgeschlossen wird, können Stichworte wie „Mädchenschule“ oder „Frauenfußball-Verein“ den Algorithmus veranlassen, der Bewerberin eine negative Bewertung zu geben. Dieser Effekt verstärkt sich durch den Einsatz der Software immer weiter, wenn unter den Angestellten immer weniger Frauen zu finden sind. Denn wird der Algorithmus auf diesen von ihm selbst erzeugten neuen Daten weiter trainiert, bestätigt sich die von ihm getroffene Auswahl scheinbar, und das Ungleichgewicht setzt sich immer weiter fort.

Das Recruiting-Projekt von Amazon wurde 2017 eingestellt.

*Von Google-Foto als „Gorilla“ bezeichnet zu werden macht Betroffene und Leser fassungslos. Eine systematische Diskriminierung im Bewerbungsprozess entscheidet über das weitere Leben der Betroffenen, vor allem wenn der Einsatz von Algorithmen im Recruiting ausgeweitet wird.*

*Um solche Effekte zu vermeiden, müssen die Trainingsdaten genau betrachtet und auf ihre Vollständigkeit überprüft werden. Eine schiefe Verteilung in den Daten kann durch gezieltes Eingreifen beseitigt werden.*

*Ein hilfreicher Ansatz ist zum Beispiel mehr Diversität in den Entwicklerteams. Dadurch können derartige Fehlertreffer schneller erkannt und vermieden werden.*

## **Selbst lernende Systeme ohne Kontrolle**

Selbst lernende Algorithmen sind heute keine Seltenheit mehr. Durch neuronale Netze können Programme selbstständig lernen und erzeugen Ergebnisse, die von ihrem Programmierer nicht mehr in allen Fällen nachvollzogen werden können oder nur sehr schwer.

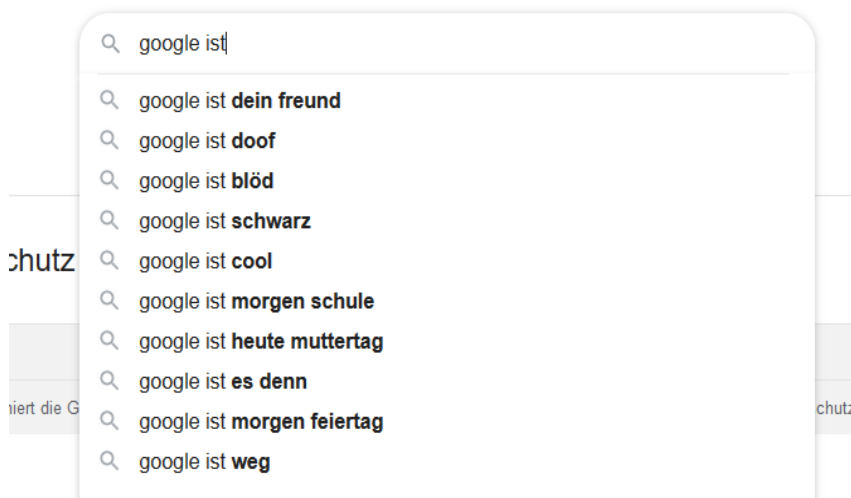
Für den Anwender sind die Vorgänge noch viel weniger zu durchschauen.

Das Internet ist voll von lustigen Assoziationen, die der Vorschlags-Algorithmus der Google-Suche hervorbringt. Doch manchmal sind die Suchvorschläge nicht lustig, sondern beleidigend oder empörend. In Deutschland ging der Prozess um Bettina Wulff durch die Medien. Wer „Bettina Wulff“ in das Suchfenster eingab, bekam „... Escort“ und weitere ähnliche Treffer geliefert. Google

beharrte darauf, dass diese Vorschläge keine eingestellten Inhalte von Google seien, sondern lediglich das Suchverhalten der Nutzer widerspiegeln. Am Ende mussten die Vorschläge jedoch entfernt werden.

In den USA kam es zu einem ähnlichen Fall. Die Google-Maps-Suche im Umkreis Washington D. C. nach den rassistischen Ausdrücken „nigger house“ bzw. „nigger king“ führten als Ergebnisse das Weiße Haus an, zu dieser Zeit das Zuhause von Präsident Obama. Google entschuldigte sich und entfernte die Treffer.

Wie die Suchvorschläge genau zu Stande kommen, ist ein Geheimnis von Google. Doch gibt das Unternehmen Auskunft über einige Details. Die Autovervollständigung macht eigentlich keine Suchvorschläge, sondern sagt die wahrscheinlichste Suche anhand der bisherigen Eingabe voraus. Diese basiert auf den Suchanfragen anderer Nutzer, sowie auf Trends (z. B. Festtage, Jahreszeiten, aktuelle Themen), dem Standort und auch der persönlichen Suchhistorie. Dadurch können sich die Vorschläge von Nutzer zu Nutzer und auch von Gerät zu Gerät unterscheiden.



Inzwischen können unangemessene Vorschläge gemeldet werden und werden dann von Google entfernt. Dennoch sieht Google sich für die Ergebnisse nicht verantwortlich und entfernt unangebrachte Verknüpfungen mit Namen nur nach Aufforderung.

Auch Facebook bedient sich einer Blackbox, um seinen Nutzern möglichst relevante Werbung und Inhalte auszuspielen. Durch die Interaktion mit Artikeln, Bannern und Seiten, aber auch die Interaktionen der Facebook-Freunde, lernt die Maschine, Inhalte auszuspielen oder zu unterdrücken. Darauf, was er sieht, hat der Nutzer selbst nur indirekt Einfluss. Facebook musste sich in diesem Zusammenhang dem Vorwurf der „Filterblase“ stellen. Nutzer bekommen immer besser zugeschnittenen Content angezeigt, wodurch sich Effekte innerhalb der Blase immer weiter verstärken. Wie weit dieser Effekt schon in der Realität angekommen ist, darüber sind sich

Experten uneins. Doch je besser der Algorithmus arbeitet, desto wahrscheinlicher kann es zu einer Filterblase kommen.

Ist für den Anwender unklar, warum er welche Inhalte zu sehen bekommt, kann der Algorithmus gezielt manipuliert werden, um bestimmte Artikel zu verbreiten, zum Beispiel durch bezahlte Freunde, die liken, teilen und kommentieren, um so eine größere Relevanz zu schaffen.

*Algorithmen müssen immer wieder von Menschen überprüft werden. Gerade wenn der Lernprozess automatisiert abläuft, müssen die Ergebnisse sorgfältig betrachtet und bewertet werden.*

*Programmierer und Unternehmen müssen Verantwortung für ihre Programme übernehmen.*

## Die Grenzen des guten Geschmacks

Bestimmte Aufgaben können Algorithmen uns nicht abnehmen. Die Software, die für Facebook über sensible Inhalte entscheidet, sorgt regelmäßig für Verwunderung und Unverständnis. Unter den gelöschten Bildern findet sich ein Foto der Neptun-Statue auf der Piazza del Nettuno in Bologna, die bis auf einen Dreizack unbedeckt ist, sowie ein bekanntes Foto aus dem Vietnamkrieg, das das Napalm Opfer Kim Phuc zeigt. Auch Bilder von stillenden Müttern verschwinden regelmäßig von der Plattform, obwohl sie den Richtlinien von Facebook nicht widersprechen.



Quelle: Wikipedia

Was Kunst oder Kulturgut ist, kann ein Computer nicht entscheiden. Nackte Haut und Brüste kennzeichnet der Facebook-Algorithmus als anstößig, einen Kontext kann er nicht bewerten.

Religion und kultureller Hintergrund lassen Menschen Bilder verschieden bewerten. Für streng Religiöse kann ein entblößter Knöchel schon unsittlich erscheinen, während andere sich vollständig entkleidet am Strand fotografieren und diese Fotos gerne in den sozialen Medien teilen wollen. Einem Algorithmus allgemein gültige Richtwerte beizubringen, kann damit nur scheitern.

Ebenso kann ein Algorithmus keine Ironie oder Satire erkennen. Der Fall von Böhmermanns Erdogan-Gedicht zeigt, dass sich auch Menschen uneinig sind, wo die Grenzen zwischen Satire und Beleidigung liegen. Da gerade überspitzte Formulierungen und auch die Überschreitung von Grenzen Kennzeichen der Satire sind, sind allgemeine Regeln nicht zu formulieren. In manchen Fällen muss ein Gericht klären, was Satire darf und was nicht mehr als Spaß gewertet werden kann. Hier sind wir auf menschliches Ermessen angewiesen, und werden es auch in Zukunft bleiben.

*Nicht jede Aufgabe lässt sich ohne weiteres in einem Programm formulieren. Ethische oder ästhetische Fragen kann eine Maschine nicht beantworten.*

Die hier betrachteten Beispiele stellen nur eine kleine Auswahl der Fälle dar, in welchen Algorithmen nicht arbeiten wie erwartet, und sich diskriminierend oder unvorhersehbar verhalten. Wer sich für weiterführende Insights zu diesem Thema interessiert, findet diese zum Beispiel in dem Buch „Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy“ von Cathy O’Neil, welches dieses Thema sehr verständlich beleuchtet und an vielen Beispielen illustriert.

Links – Stand August 2019

<https://www.spiegel.de/netzwelt/web/google-fotos-bezeichnet-schwarze-als-gorillas-a-1041693.html>

<https://www.zeit.de/digital/internet/2015-07/google-fotos-algorithmus-rassismus>

<https://www.spiegel.de/netzwelt/web/wenn-algorithmen-ueber-bewerbungen-entscheiden-a-1273429.html>

<https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/der-maschinelle-weg-zum-passenden-personal/>

<https://t3n.de/news/diskriminierung-deshalb-platzte-amazons-traum-vom-ki-gestuetzten-recruiting-1117076/>

<https://www.zeit.de/digital/internet/2012-09/reaktionen-wulff-google>

<https://www.sueddeutsche.de/medien/filterblase-facebook-youtube-soziale-netzwerke-1.4245243>

<https://twitter.com/dannysullivan/status/966610020040241152>

<https://www.blog.google/products/search/how-google-autocomplete-works-search/>

<https://www.theguardian.com/technology/2015/may/20/google-apologises-racist-google-maps-white-house-search-results>

<https://www.faz.net/aktuell/feuilleton/google-entfernt-ergaenzungen-bei-suche-nach-bettina-wulff-13373712.html>

<https://www.spiegel.de/netzwelt/web/facebook-sperrt-anzeige-mit-bild-von-nacktem-neptun-a-1128337.html>

<https://www.spiegel.de/netzwelt/web/facebook-loeschungen-bildet-eure-mitarbeiter-besser-aus-kommentar-a-1111655.html>