



Gert G. Wagner, Senior Research Fellow  
am DIW Berlin

Der Kommentar gibt die Meinung des Autors wieder.

GERT G. WAGNER

# Schlagwort „Open Source“: Algorithmen können und müssen anhand von Beispielfällen transparent gemacht werden

Man muss nicht mal besonders schreckhaft sein, um vor „selbstlernenden Computeralgorithmen“, die etwa die Kreditwürdigkeit eines Einzelnen bestimmen, Angst zu haben. Auch nüchterne Betriebswirte oder Ingenieure fürchten sich davor, dass „Künstliche Intelligenz“ sie nicht nur überflüssig macht, sondern auch, dass Computer, die sich selbst was Neues beibringen, die Macht übernehmen. Die Maschinen sind allerdings bei weitem noch nicht so schlau, und verstehen kann man sie auch. Man darf freilich nicht jede Schlagzeile über Big Data blind vertrauen.

Zuerst einmal sei festgehalten: Nicht nur Computeralgorithmen, die ständig mit neuen Daten gefüttert werden, und so „selbst lernen“, sind schwer zu begreifen. Dasselbe gilt für seit Jahrzehnte benutzte lineare Regressionsmodelle, wenn diese Interaktionsterme (also die Multiplikation zweier Variablen wie Einkommen und Alter) oder Potenzen und Polynome enthalten (also etwa neben der Berufserfahrung in Jahren auch deren Quadrat). Die Wirkungen solcher Regressionsgleichungen werden daher auch in wissenschaftlichen Veröffentlichungen zunehmend anhand von Beispielen und Graphiken anschaulich gemacht.

Algorithmen auf der Basis sogenannter „neuronalen Netze“, also nichtlinearer „Lernmaschinen“, sind noch schwerer zu verstehen – selbst für diejenigen, die sie programmieren. Obwohl diese den Programmcode kennen, ist es ihnen nicht ohne weiteres möglich, zu verstehen, was das künstliche Gehirn mit den Daten wirklich macht.

Für einzelne Entscheidungssituationen (etwa die Bewertung der Kreditwürdigkeit einer bestimmten Person) kann eine algorithmische Entscheidung im Nachhinein erklärt werden – das bietet zum Beispiel die Webseite [www.heatmaping.org](http://www.heatmaping.org) an. Dabei werden Beispielfälle durch das neuronale Netz geschickt. Durch den Vergleich der Ergebnisse für verschiedene Beispielfälle wird die Offenlegung der Mechanismen prinzipiell möglich. In der Wissenschaft wird inzwischen von „erklärbarer Künstlicher Intelligenz“ gesprochen, der dazu notwendige technische Aufwand dürfte rasch kleiner werden. Dabei geht es nicht darum, jedes künstliche Neuron in einem neuronalen Netz zu beob-

achten oder gar zu verstehen. Man will lediglich wissen und veranschaulichen können, zu welchen Entscheidungen ein Netz – oder herkömmliche Regressionsmodelle, wie sie etwa die Schufa nutzt – bei bestimmten Konstellationen kommt („black box tinkering“). Dies kann auch anhand von systematisch ausgewählten Beispielprofilen erfolgen. Damit kann man im Prinzip auch herausfinden, ob ein Algorithmus direkt oder indirekt bestimmte Personengruppen diskriminiert. Es könnte zum Beispiel sein, dass der Kreditwürdigkeit-Score von Menschen, die zwar immer ihre Rechnung pünktlich bezahlen, aber unzuverlässige Nachbarn haben, niedrig ist und dies als Diskriminierung bewertet wird.

Um die Algorithmen „erklärbar“ zu machen, müssen ihre Hersteller verpflichtet werden, die benutzten Merkmale offenzulegen (etwa Geschlecht, Alter, Art der Internetnutzung) – der Computercode selbst bleibt ihr Geheimnis. Am besten wäre es, eine Plattform einzurichten, etwa von einer „Digitalagentur“, wie sie der Sachverständigenrat für Verbraucherfragen vorschlägt. Hier könnte jedermann auch seine eigenen Merkmale eingeben, um die Wirkung eines Algorithmus zu simulieren.

Die Europäische Datenschutzverordnung bietet Anknüpfungspunkte, etwa mit der Forderung, dass Menschen Algorithmen prüfen können müssen, wenn diese ihre „berechtigten Interessen (...) erheblich beeinträchtigen“. Und im Koalitionsvertrag der Bundesregierung ist der Wille, die derzeitige Intransparenz zu überwinden, explizit festgeschrieben. Dazu soll eine Ethikkommission eingesetzt werden soll. Diese wird sich nicht hinter dem Geschäftsgeheimnis verstecken können: Das Durchrechnen von Beispielfällen durch eine zur Verschwiegenheit verpflichtete Aufsichtsbehörde verletzt dieses auf keinen Fall.

Dieser Beitrag beruht auf einem in Zusammenarbeit mit Gerd Gigerenzer und Klaus-Robert Müller entstanden und am 22. Juni 2018 in der Frankfurter Allgemeinen Zeitung erschienenen Artikel.

## IMPRESSUM

---



DIW Berlin — Deutsches Institut für Wirtschaftsforschung e.V.

Mohrenstraße 58, 10117 Berlin

[www.diw.de](http://www.diw.de)

Telefon: +49 30 897 89-0 Fax: -200

85. Jahrgang 27. Juni 2018

### Herausgeberinnen und Herausgeber

Prof. Dr. Tomaso Duso; Dr. Ferdinand Fichtner; Prof. Marcel Fratzscher, Ph.D.;  
Prof. Dr. Peter Haan; Prof. Dr. Claudia Kemfert; Prof. Dr. Alexander Kriwoluzky;  
Prof. Dr. Stefan Liebig; Prof. Dr. Lukas Menkhoff; Prof. Johanna Möllerström,  
Ph.D.; Prof. Karsten Neuhoff, Ph.D.; Prof. Dr. Jürgen Schupp;  
Prof. Dr. C. Katharina Spieß

### Chefredaktion

Dr. Gritje Hartmann; Mathilde Richter; Dr. Wolf-Peter Schill

### Lektorat

Manja Gärtner, Ph.D.; Dr. Dietmar Edler

### Redaktion

Renate Bogdanovic; Dr. Franziska Bremus; Rebecca Buhner;  
Claudia Cohnen-Beck; Dr. Daniel Kemptner; Sebastian Kollmann;  
Matthias Laugwitz; Markus Reiniger; Dr. Alexander Zerrahn

### Vertrieb

DIW Berlin Leserservice, Postfach 74, 77649 Offenburg

[leserservice@diw.de](mailto:leserservice@diw.de)

Telefon: +49 1806 14 00 50 25 (20 Cent pro Anruf)

### Gestaltung

Roman Wilhelm, DIW Berlin

### Umschlagmotiv

© imageBROKER / Steffen Diemer

### Satz

Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

### Druck

USE gGmbH, Berlin

ISSN 0012-1304; ISSN 1860-8787 (online)

Nachdruck und sonstige Verbreitung – auch auszugsweise – nur mit  
Quellenangabe und unter Zusendung eines Belegexemplars an den  
Kundenservice des DIW Berlin zulässig ([kundenservice@diw.de](mailto:kundenservice@diw.de)).

Abonnieren Sie auch unseren DIW- und/oder Wochenbericht-Newsletter  
unter [www.diw.de/newsletter](http://www.diw.de/newsletter)