



Sachstand

Aspekte zu Künstlicher Intelligenz und Black-Box-Systemen

Aspekte zu Künstlicher Intelligenz und Black-Box-Systemen

Aktenzeichen: WD 8 - 3000 - 063/19
Abschluss der Arbeit: 30.04.2019
Fachbereich: WD 8: Umwelt, Naturschutz, Reaktorsicherheit, Bildung und
Forschung

Die Wissenschaftlichen Dienste des Deutschen Bundestages unterstützen die Mitglieder des Deutschen Bundestages bei ihrer mandatsbezogenen Tätigkeit. Ihre Arbeiten geben nicht die Auffassung des Deutschen Bundestages, eines seiner Organe oder der Bundestagsverwaltung wieder. Vielmehr liegen sie in der fachlichen Verantwortung der Verfasserinnen und Verfasser sowie der Fachbereichsleitung. Arbeiten der Wissenschaftlichen Dienste geben nur den zum Zeitpunkt der Erstellung des Textes aktuellen Stand wieder und stellen eine individuelle Auftragsarbeit für einen Abgeordneten des Bundestages dar. Die Arbeiten können der Geheimschutzordnung des Bundestages unterliegende, geschützte oder andere nicht zur Veröffentlichung geeignete Informationen enthalten. Eine beabsichtigte Weitergabe oder Veröffentlichung ist vorab dem jeweiligen Fachbereich anzuzeigen und nur mit Angabe der Quelle zulässig. Der Fachbereich berät über die dabei zu berücksichtigenden Fragen.

Inhaltsverzeichnis

1.	Einleitung	4
2.	Transparenz und Kontrolle von Algorithmen	4
3.	Fazit	6
4.	Quellenverzeichnis	7

1. Einleitung

Ein Algorithmus ist ein eindeutig geregeltes Verfahren zur Lösung einer Aufgabe und besteht aus einer Reihe von Anweisungen, deren einzelne Schritte definiert sind. Die Anweisungen erfolgen seriell, können aber auch verzweigen. Im Idealfall gibt es keinen Spielraum für Interpretationen. Der Algorithmus darf nur eine Anzahl von endlichen Schritten benötigen und nicht in eine Endlosschleife geraten. Grundlage für immer komplexere und schnellere Algorithmen bietet eine wachsende Datenbasis (Big Data). Algorithmen sind in der Regel nicht öffentlich verfügbar, und aufgrund der fehlenden Transparenz ist eine Einordnung nach z.B. ihrer Verarbeitungslogik, Wirkung oder Funktionsweise nicht möglich. Computerprogramme setzen Algorithmen um und gelten als urheberrechtlich geschütztes Werk, wogegen reine Rechenregeln, wie Algorithmen, als nicht schutzfähig gelten.¹

Algorithmische Entscheidungssysteme (algorithmic decision making systems oder auch ADM-Systeme) beinhalten Regeln, nach denen eine Entscheidung getroffen werden kann. Zudem gibt es Entscheidungssysteme deren Entscheidungsregeln von Algorithmen selbständig abgeleitet werden, wie z.B. beim Deep Learning. Nicht immer sind diese Entscheidungen transparent nachvollziehbar und die Forscher sprechen von „Black-Box“-Systemen.

Die vorliegende Arbeit befasst sich im Rahmen der Algorithmus-Kontrolle mit Aspekten zum Stand der Umsetzbarkeit.

2. Transparenz und Kontrolle von Algorithmen

Mit Zunahme der Verwendung von Algorithmen bzw. ADM-Systemen wird auch die Forderung laut, dass die Algorithmen transparent nachvollziehbar und kontrollierbar sein müssen. Zumindest wenn ihre Entscheidungen einzelne Menschen oder die Gesellschaft als Ganzes betreffen. Mit Algorithmus unterstützter Entscheidungsfindung kann ein hoher Schaden nur für einen einzelnen oder für die gesamte Gesellschaft entstehen.

1 Wissenschaftliche Dienste (2017). Aktueller Begriff „Einsatz und Einfluss von Algorithmen auf das digitale Leben“, <https://www.bundestag.de/resource/blob/530808/b07930558a906b8ad51a9cde7ef8f11d/algorithmen-data.pdf>

Wissenschaftliche Dienste (2017). Sachstand „Algorithmen Einzelfragen zu Instrumenten und Regelansätzen“, <https://www.bundestag.de/resource/blob/529290/6c67be680882ef8e04fa752ab2a15c34/wd-8-031-17-pdf-data.pdf>

Rankin, K., MIT Technology Review (2019). „The Dark Secret at the Heart of AI“, <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>

MIT Computer Science & Artificial Intelligence Lab (2016). „Making computers explain themselves“, <https://www.csail.mit.edu/news/making-computers-explain-themselves>

Ob ADM-Systeme überprüfbar sind, hängt auch von ihrer Komplexität ab. Die künstlichen neuronalen Netze des Deep Learnings beispielsweise gelten als nicht-erklärend.² Informationen werden in den neuronalen Netzen nicht an einem bestimmten Ort gespeichert, sondern sind in einer Vielzahl von Verbindungen abgelegt, die teilweise nicht überschaubar sind. Kritiker fordern eine Kontrolle dieser Black-Box-Systeme der Künstlichen Intelligenz (KI). Gegner der Kontrolle schätzen die Komplexität der KI-Systeme: „Ohne die komplexen Antworten des Maschinenlernens würde im wissenschaftlichen Werkzeugkasten ein wichtiges Instrument fehlen, argumentieren die Forscher. Denn die reale Welt sei nun einmal komplex: Für manche Phänomene wie das Wetter oder den Aktienmarkt existiert vielleicht nicht einmal eine reduktionistische und synthetische Beschreibung.“ Zudem verarbeiten selbstlernende Systeme Datenmengen, die ohne sie nur schwer zu bearbeiten wären.³

Um Forschung und Entwicklung nicht zu beschneiden, gleichzeitig aber nicht vor einem nicht nachvollziehbarem Ergebnis zu stehen und möglicherweise der Gesellschaft oder dem Individuum Schaden zuzuführen, ist auch zu überlegen, ob in gewissen sensiblen Bereichen, wie beispielsweise beim „Predictive Policing“, der Berechnung der Rückfallwahrscheinlichkeit von Straftätern, oder bei der Krebsvorsorge, selbstlernende Algorithmen, deren Ergebnis bzw. Erkenntnisweg nicht transparent ist, nicht zuzulassen, oder deren Ergebnis menschlichen Erwägungen beratend oder gewichtend zur Seite zu stellen.⁴

Verschiedene theoretische Ansätze befassen sich mit der Transparenz und Rechenschaftspflicht (Accountability) von insbesondere selbstlernenden Algorithmen.⁵

2 Konrad Adenauer Stiftung (2019). „Algorithmische Entscheidungen: Transparenz und Kontrolle“, Analysen und Argumente Nr. 338, <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>

3 Spektrum der Wissenschaft (2016). „Eine tückische Blackbox“, <https://www.spektrum.de/news/eine-tueckische-blackbox/1429906> vom 16.11.2016

4 Konrad Adenauer Stiftung (2019). „Algorithmische Entscheidungen: Transparenz und Kontrolle“, Analysen und Argumente Nr. 338, <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>

Spektrum der Wissenschaft (2016). „Eine tückische Blackbox“, <https://www.spektrum.de/news/eine-tueckische-blackbox/1429906> vom 16.11.2016

5 Pavlus, J., Quantamagazine (2019). „A New Approach to Understanding How Machines Think“, <https://www.quantamagazine.org/been-kim-is-building-a-translator-for-artificial-intelligence-20190110/> vom 10.1.2019

Wolchover, N., Quantamagazine (2017). „New Theory Cracks Open the Black Box of Deep Learning“, <https://www.quantamagazine.org/new-theory-cracks-open-the-black-box-of-deep-learning-20170921/> vom 21.9.2017

Fraunhofer IUK-Technologie (2017). „Blick in die Black Box“, <https://www.fraunhofer-innovisions.de/ce-bit/blick-in-die-black-box> vom 9.3.2017

In der Regel können selbstlernende Algorithmen ihre berechneten Entscheidungen nicht begründen. In Bereichen, in denen das notwendig ist, können nur transparent nachvollziehbare Ergebnisse verwendet werden („rights to explain“).⁶

Auch ist die Problematik des "foolings", bei der Algorithmen getäuscht und manipuliert, bzw. gehackt werden können nicht gelöst.⁷

Im Rahmen der Wahrung der Privatsphäre wäre beispielsweise zu überlegen, ob die Ansätze des Digital Literacy z.B. in Schulen verankert werden und damit den Einzelnen darüber aufzuklären, inwieweit er sich beispielsweise in den sozialen Netzwerken den Risiken der Algorithmen aussetzt.

3. Fazit

Die Ergebnisse der Algorithmen gelten insbesondere im Bereich der Künstlichen Intelligenz nicht immer als transparent. Forderungen nach Transparenz und Rechenschaftspflicht gibt es schon jetzt. Ob eine generelle Kontrollmöglichkeit gegeben sein muss oder diese nur für ausgewählte Anwendungsfelder eingeführt werden sollte, ist in der offenen Diskussion.

Fraunhofer IUK-Technologie (2017). „CeBIT 2017: Analyse-Software für neuronale Netze - Dem Computer beim Denken zuschauen“, <https://www.hhi.fraunhofer.de/presse-medien/nachrichten/archiv/2017/2017/cebit-2017-analyse-software-fuer-neuronale-netze-dem-computer-beim-denken-zuschauen.html> vom 8.2.2017

6 Konrad Adenauer Stiftung (2019). „Algorithmische Entscheidungen: Transparenz und Kontrolle“, Analysen und Argumente Nr. 338, <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>

Spektrum der Wissenschaft (2016). „Eine tückische Blackbox“, <https://www.spektrum.de/news/eine-tueckische-blackbox/1429906> vom 16.11.2016

7 Spektrum der Wissenschaft (2016). „Eine tückische Blackbox“, <https://www.spektrum.de/news/eine-tueckische-blackbox/1429906> vom 16.11.2016

4. Quellenverzeichnis

Fraunhofer IUK-Technologie (2017). „Blick in die Black Box“, <https://www.fraunhofer-innovations.de/cebit/blick-in-die-black-box> vom 9.3.2017

Fraunhofer IUK-Technologie (2017). „CeBIT 2017: Analyse-Software für neuronale Netze - Dem Computer beim Denken zuschauen“, <https://www.hhi.fraunhofer.de/presse-medien/nachrichten/archiv/2017/2017/cebit-2017-analyse-software-fuer-neuronale-netze-dem-computer-beim-denken-zuschauen.html> vom 8.2.2017

Konrad Adenauer Stiftung (2019). „Algorithmische Entscheidungen: Transparenz und Kontrolle“, Analysen und Argumente Nr. 338, <https://www.kas.de/documents/252038/4521287/AA338+Algorithmische+Entscheidungen.pdf/533ef913-e567-987d-54c3-1906395cdb81?version=1.0&t=1548228380797>

MIT Computer Science & Artificial Intelligence Lab (2016). „Making computers explain themselves“, <https://www.csail.mit.edu/news/making-computers-explain-themselves>

Pavlus, J., Quantamagazine (2019). „A New Approach to Understanding How Machines Think“, <https://www.quantamagazine.org/been-kim-is-building-a-translator-for-artificial-intelligence-20190110/> vom 10.1.2019

Rankin, K., MIT Technology Review (2019). „The Dark Secret at the Heart of AI“, <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>

Spektrum der Wissenschaft (2016). „Eine tückische Blackbox“, <https://www.spektrum.de/news/eine-tueckische-blackbox/1429906> vom 16.11.2016

Wissenschaftliche Dienste (2017). Aktueller Begriff „Einsatz und Einfluss von Algorithmen auf das digitale Leben“, <https://www.bundestag.de/resource/blob/530808/b07930558a906b8ad51a9cde7ef8f11d/algorithmen-data.pdf>

Wissenschaftliche Dienste (2017). Sachstand „Algorithmen Einzelfragen zu Instrumenten und Regelansätzen“, <https://www.bundestag.de/resource/blob/529290/6c67be680882ef8e04fa752ab2a15c34/wd-8-031-17-pdf-data.pdf>

Wolchover, N., Quantamagazine (2017). „New Theory Cracks Open the Black Box of Deep Learning“, <https://www.quantamagazine.org/new-theory-cracks-open-the-black-box-of-deep-learning-20170921/> vom 21.9.2017