

Andreas Sudmann

Deep Learning als Black
Box, *OpenAI* und das
Internet



CAIS Report

Fellowship
Mai bis Oktober 2017

Deep Learning als Black Box, *OpenAI* und das Internet

Zur medienpolitischen Dimension maschineller Lernverfahren

Deep Learning zwischen Black-Box-Technologie und demokratischen Ansprüchen

Die internetbasierte Digitalkultur wird zunehmend durch moderne Technologien der Künstlichen Intelligenz (KI) bestimmt. Vor allem Ansätze des sogenannten Deep Learning (DL), die sich an der Funktionsweise biologischer Gehirne orientieren, haben derzeit Konjunktur. Als maschinelle Lernverfahren ermöglichen sie es, dass ein Computer lernen kann, ohne für diesen Zweck explizit programmiert zu werden.

Das Forschungsvorhaben analysiert die medienpolitische Dimension maschineller Lernverfahren. Zwei Aspekte stehen dabei im Zentrum: Zum einen geht es um die Frage, inwieweit Deep Learning eine Black-Box-Technologie darstellt, d.h. als System zu betrachten ist, dessen interne Operationen und Funktionen dem menschlichen Verständnis versperrt sind.

Zum anderen sollen industriell-wissenschaftliche Vorhaben (z.B. das Non-Profit-Forschungsunternehmen *OpenAI*) untersucht werden, die zumindest dem eigenen Anspruch nach um eine Demo-

kratisierung der Künstlichen Intelligenz bemüht sind.

Das Projekt prüft dabei nicht nur, ob die Ansprüche eingelöst werden. Vielmehr analysiert es die unterschiedlichen medialen Bedingungen der maschinellen Lernverfahren und verortet sie theoretisch, historisch und politisch.

Zwei Analyseperspektiven im Fokus

Das allgemeine und übergeordnete Ziel des vorliegenden Forschungsvorhabens bestand darin, das Feld wechselseitiger Relationen, Funktionen sowie potentieller Effekte zwischen dem Internet auf der einen und DL-Technologien auf der anderen Seite in medienpolitischer Hinsicht zu vermessen. Dabei standen zwei Forschungsperspektiven/-aspekte im Zentrum der Studie:

- a) der generelle Status internetbasierter DL-Verfahren als vermeintliche Black Box-Technologie und seine spezifischen medienpolitischen Implikationen, Funktionen und Effekte,
- b) medienpolitische Unternehmen/Projekte der internetgestützten Demokratisierung von KI-/DL-

Technologien, mit Schwerpunkt auf das Unternehmen OpenAI als Fallstudie.

Beide Analyseperspektiven sind eng miteinander verzahnt: Wie kann ein finanzstarkes, vermeintliches Non-Profit-Unternehmen wie OpenAI zur Transparenz von DL-Verfahren beitragen, wenn letztere sich dadurch auszeichnen, fundamental undurchsichtig bzw. gar selbst für menschliche IT-Expert*innen und ihre Entwickler*innen unverständlich zu sein? Umgekehrt scheint gerade die unterstellte Undurchsichtigkeit des DL eine der Möglichkeitsbedingungen dafür zu sein, dass sich Projekte wie OpenAI gesellschaftlich und öffentlichkeitswirksam legitimieren können, ganz unabhängig von der Frage, inwieweit Unternehmen wie OpenAI letztlich auch durch ökonomische Interessen motiviert sind.

Ein wesentlicher Ertrag des Projekts besteht darin, ein genaueres Verständnis über die medialen und medienpolitisch relevanten Bedingungen der internetbasierten Digitalkultur zu gewinnen und auf diese Weise die sozialen Chancen, Risiken und Problemstellungen der Verbindung von Internet- und KI-Technologien kompetenter beschreiben und historisch-theoretisch erklären zu können.

Eine Ausgangsthese des Projekts lautete dabei, dass DL-Verfahren und Internet-Technologien als soziotechnische Praxis funktional wechselseitig aufeinander bezogen sind: DL-Verfahren ermöglichen nicht nur eine „smarte“ Version von Internet-Dienstleistungen und -technologien, vielmehr sind sie wesentlich von einem internetbasierten Zugriff auf große, vernetzte Datenmengen abhängig, mit jeweils sehr relevanten politischen Implikationen und schon jetzt absehbaren Effekten (z.B. Umgestaltungen von Arbeitsprozessen etc.). Gleichzeitig sind speziell in jüngerer Zeit DL-Forscher*innen daran interessiert, diese Abhängigkeit von großen Datenmengen und der aufwendigen internetbasierten Suche nach geeigneten Daten zu reduzieren, zum Beispiel durch Verfahren des sogenannten Transfer Learnings.

Theorien, Gegenstand und Methoden

In theoretischer Hinsicht bezieht sich die Studie u.a. auf Ansätze der sogenannten Akteur-Netzwerk-Theorie (ANT), die seit den 1980er Jahren durch Personen wie Bruno Latour, John Law und

Michel Callon im Feld der Science and Technology Studies (STS) prominent vertreten wird. Ein prominentes Kennzeichen der ANT besteht darin, Strukturen verteilter Handlungsmacht zwischen Dingen und Menschen aufzuzeigen sowie das Geflecht unterschiedlicher Vermittlungsgrößen (Hardware, Software, KI-Labore, IT-Konzerne etc.) in der Hervorbringung und Formierung von Technologien sichtbar zu machen.

Der Fokus auf die medienpolitische Dimension der KI bedeutet also nicht, die politischen Aspekte der massenmedialen Darstellung von KI-Verfahren zu untersuchen, sondern die vernetzten Konfigurationen und Konstellationen der Technologien selbst hinsichtlich ihrer politischen Implikationen zu befragen.

Die geplante Studie wertet in erster Linie das konsolidierte Wissen von KI-Expert*innen in Form von Interviews, wissenschaftlichen Veröffentlichungen, Presseberichten etc. aus. Zur Eingrenzung der Forschungsperspektive steht dabei Material im Zentrum, das sich auf maschinelle Lernaufgaben der Bild- und Texterkennung bezieht.

Methodisch greift das Projekt u.a. auf Ansätze der Diskursanalyse als qualitatives Verfahren der Medienforschung zurück. Die Methode soll vor allem dazu dienen, den Selbstbeschreibungen von Deep-Learning-Expert*innen kritisch begegnen zu können. Dadurch versucht sie vor allem das politische Regime des Sagbaren und Unsagbaren auszuloten.

Maschinelle Lernverfahren sind politisch nicht neutral

Eine zentrale These als Ergebnis des Projekts lautet, dass die maschinellen Lernverfahren des Deep Learning unabhängig von ihrer jeweils konkreten Funktion (z.B. automatische Erkennungsverfahren terroristischer Aktivitäten, angewandt etwa bei E-Mail-Kommunikation, Auswertung von Social Media-Inhalten für Wahlkampfzwecke) politisch relevant sind.

Maschinelle Lernverfahren sind also grundsätzlich keine neutrale Technologie, die gleichsam in einem politikfreien Raum existiert. Was Algorithmen überhaupt lernen können, wird wesentlich durch menschliche Entwickler*innen bestimmt, zum Beispiel durch die Auswahl des entsprechenden

Datenmaterials, das die Grundlage maschineller Lernprozesse bildet.

Insofern sind Vorstellungen einer fundamentalen Autonomie dieser Lerntechnologie grundsätzlich falsch. Soziale Normen, Werte und spezifische kulturelle Konstellationen beeinflussen die technologischen Bedingungen maschineller Lernverfahren, ohne dass man diesen Einfluss den Verfahren ohne Weiteres ablesen könnte.

OpenAI löst das Versprechen einer Demokratisierung von KI nur bedingt ein

Ein weiteres Resultat lautet, dass Labore wie OpenAI derzeit weniger durch ihre Arbeit an einer Demokratisierung der KI auffallen. Das Unternehmen versucht eher, auf ethisch verantwortliche Weise die Grundlagenforschung im Bereich des DL voranzutreiben.

In Bezug auf viele Projekte, die OpenAI verfolgt, ist der Zusammenhang mit einer politisch ambitionierten KI-Agenda nur bedingt erkennbar. Auch die technologischen Entwicklungen kommen nicht wirklich dadurch zur Abhebung, dass sie mehr als andere Unternehmen tatsächlich „transparenter“ sind.

Ferner sind Unterschiede zwischen OpenAI und der KI-Politik von Unternehmen wie Microsoft oder Facebook nur bedingt erkennbar. Als Grund hierfür wurde u.a. ermittelt, dass die genannten Konzerne, z.T. in Reaktion auf OpenAI, eine ähnliche KI-Politik verfolgen, insofern sie ebenfalls offiziell um die Entwicklung einer transparenten, sozial verantwortlichen KI bemüht sind. Gleichwohl ist auch bei IT-Unternehmen wie Facebook festzustellen, dass Anspruch und Wirklichkeit stark auseinandergehen.

Detaillierte Ergebnisse finden sich unter anderem bei Sudmann (2018).

Literaturverzeichnis

Ruder, Sebastian (2017). *Transfer Learning – Machine Learning's Next Frontier*. Verfügbar unter <https://ruder.io/transfer-learning/> [17.01.2020].

Sudmann, Andreas (2018). On the Media-political Dimension of Artificial Intelligence: Deep Learning as a Black Box and OpenAI. *Digital Culture & Society* 4(1), 181–200.

DOI: <https://doi.org/10.14361/dcs-2018-0111>

Abbildungsverzeichnis

Foto Titelseite: CAIS, Matthias Begenat

Kontakt

PD Dr. Andreas Sudmann
Ruhr-Universität Bochum
Email: sudmann@rub.de