

# Die Magie von KI: Algorithmen dringen in die Arbeitswelt

Bericht von den 7. DGI-Praxistagen unter dem Motto: „Der Algorithmus wo man mitmuss“ über neue Wege bei der Informationserschließung, Frankfurt am Main, 8. und 9. November 2018

Elgin Helen Jakisch

Die 7. DGI-Praxistage widmeten sich der Frage, inwieweit Technologien der Künstlichen Intelligenz (KI), des Data Mining und maschinellen Lernens die Arbeitswelt der Information Professionals verändern oder ob sie diese gar ersetzen. Experten aus Forschung und Praxis berichteten von ihren Projekten und Erfahrungen beim Einsatz von Algorithmen zur Erschließung von Medien aller Art. „Wie funktionieren die Werkzeuge und wo ist menschliche Intelligenz unabdingbar?“ fragte Marlies Ockenfeld, Präsidentin der Deutschen Gesellschaft für Information und Wissen (DGI), bei der Begrüßung der Teilnehmenden. Genau diesem Spannungsfeld ging die Tagung nach.

› Mit aktuellen Möglichkeiten und Grenzen der KI setzte sich Prof. Harald Sack (KIT und FIZ Karlsruhe) in seiner Eröffnungsk keynote anhand der historischen Entwicklung der Technologie auseinander. Er nannte Beispiele faszinierender Möglichkeiten von Algorithmen, die Fotos zu Gemälden im Stil von Monet umwandeln oder KI, die Musik „wie Bach“ komponiert bis hin zu dem berühmten Computer, der den Menschen bei Go-Spiel geschlagen hat. Harald Sack sieht auch die Grenzen dieser Technologie. Bisherige Versuche, den Menschen komplett zu ersetzen, seien immer gescheitert.

Die Magie von KI, so Sack, rühre letztlich aus der Möglichkeit, sowohl orientiert an der menschlichen Intelligenz als auch an Mustererkennungen in einer enorm hohen Geschwindigkeit „genaue, zufriedenstellende oder sogar großartige“ Ergebnisse zu liefern. Der Traum der KI-Entwickler sei der gleiche, den einst Leibniz hatte, nämlich Schlussfolgerungen berechenbar zu machen. Soweit sei es noch nicht, da implizites und Allgemeinwissen nach wie vor schwer kategorisierbar seien. „Strukturierte Fakten sind für Algorithmen bisher am einfachsten zu verarbeiten“, erklärte Sack. Langfristig würden sich eine implizite und explizite Semantik entwickeln, die „Entscheidungsinformationen vorbereiten werden, die uns auch gesellschaftspolitisch betreffen“. Hierzu bedürfe es dringend einer Ethik und es brauche Transparenz und Regulierungen, empfahl er in seiner Keynote.

## Die menschliche Komponente

Gar nicht magisch, sondern ganz konkret wird seit ei-



nigen Jahren erfolgreich KI zur maschinellen Klassifikation von bibliographischen Daten bei der Deutschen Nationalbibliothek (DNB) eingesetzt. Der Rückgang an Printmedien, der Zuwachs an Online-Medien und die technischen Möglichkeiten, mit Hilfe der hauseigenen Normdateien und DDC-Sachgruppen zu arbeiten, veranlasste die DNB, diesen Weg zu gehen. Frank Busse erläuterte den Workflow, wie bei der Nationalbibliothek Netzpublikationen automatisch inhaltlich erschlossen werden. Dabei handle es sich um „unkörperliche Medien“ wie E-Books, E-Journals oder Websites mit Textinhalten. Trainingskorpora werden aus Inhaltsverzeichnissen erstellt und einer linguistischen Analyse unterworfen, die Schlagwort-Vorschläge macht. Busse erklärte, hierbei müsse der Mensch durch Monitoring beim Qualitätsmanagement an der Verbesserung der Verfahren mitwirken. Dabei sei re-

*Harald Sack und die Faszination der Magie von KI.*

levant, ob die maschinellen mit den intellektuellen Ergebnissen übereinstimmen. Bisher sei das bei 76% der Fall und „das könnte verbessert werden“, so Busse. Bei der Schlagwort-Auswahl werde im Katalog gekennzeichnet, ob ein Mensch oder eine Maschine eine Publikation inhaltlich erschlossen hat. Über 1,7 Mio. Datensätze sind bereits auf diese Weise erschlossen. Auch ZBW – Leibniz Informationszentrum Wirtschaft erprobt neue KI-gestützte Verfahren. Dr. Andreas Oskar Kempf (ZBW, Kiel) zeigte Auswirkungen auf die Arbeit des Fachreferenten



*Andreas Mertens und Interessierte schauen dem Algorithmus von avameo beim Arbeiten über die Schulter.*

beim Thesaurusmanagement. Der Auftrag an die ZBW, anwendungsorientierte Miningverfahren für das Bibliothekswesen zu sondieren, waren der Anlass für die ZBW, ihre Prozesse bei der Weiterentwicklung des Thesaurus zu optimieren. Das Ziel sei unter anderem eine internationale Vereinheitlichung terminologischer Normierungen. Das Problem, so Kempf, seien „Synonyme und Quasi-Synonyme, Eindeutigkeiten der Begriffe und die Anschlussfähigkeit zu anderem Vokabular“. Ebenso müssten Crosskonkordanzen hinterlegt werden wie der Standard-Thesaurus-Wirtschaft oder die Gemeinsame Normdatei GND. Die ZBW sei gerade dabei, Assoziationsrelationen zu Wikimedia herzustellen, berichtete Kempf. „Der Umgang mit sprachlichem Wandel ist eine stetige Herausforderung“, hob er hervor. Ein Redaktionsgremium müsse letztlich entscheiden. Kempf erläuterte die neuen Aufgaben der Fachreferenten. Diese arbeiteten jetzt an Textkorpora, Trainingsdaten, Regeln, Revisionen und an der Bewertung der KI-Verfahren. Der Mensch müsse auch hier wie beim Beispiel der DNB die Vorschläge der Maschine gewichten. Sein Fazit: „Man muss die Auswirkungen des Einsatzes von KI immer in einen größeren Zusammenhang stellen, weil konkrete Arbeitsaufgaben verändert werden.“<sup>1</sup>

Auch die Deutsche Bundesbank setzt maschinelle Lernverfahren, Machine Learning, ein. Hendrik Doll vom Forschungsdaten- und Servicezentrum FDSZ erläuterte ein automatisiertes Verfahren, mit dem neue Erkenntnisse aus Forschungsartikeln aus Open-Science-Quellen extrahiert werden. Mit Hilfe von vergleichendem Vorfiltern nach Ähnlichkeiten aus Textkörpern würden auch hier Zweifelsfälle für Sachbe-

arbeiter herausgefiltert, die manuell nachbearbeitet werden müssen, letztlich aber eine Zeitersparnis darstellen, weil insgesamt eine große Menge an Informationen verarbeitet wird. Die Bundesbank kombiniert ihre Verfahren mit eigenen Daten und Daten von Providern (u.a. Bureau van Dijk). Auch Reviews über Texte werden einbezogen, ähnlich wie bei Amazon oder Trip Advisor.

### KI-Entwickler sind „Legobauer“

Wie entstehen automatische Textzusammenfassungen? Die Firma avameo, Wiesbaden, bietet „Automatic Summarizing“ für Firmenkunden. „Extraktives und abstraktives Zusammenfassen wird bei avameo mit neurolinguistischem Programmieren kombiniert“, erläuterte Andreas Mertens, Inhaber von avameo. Als Basis dienen Taxonomien, Ontologien und Klassifikationen. Genutzt würden auch Open-Source-basierte Algorithmen oder Informationen aus Patentanmeldungen und offen zugänglichen Quellen. Metadaten würden dann aus Überschriften gewonnen. Texte mit einem hohen Qualitätsstandard wie wissenschaftliche Artikel würden bevorzugt. „Der Algorithmus lernt ständig dazu“, so Mertens. Für den Kunden würden Extrakte, Quellenangaben, Zitate und Referenzen mit einer Verlinkung zum Original versehen. „Ich glaube nicht, dass es DIE eine Methode zur Zusammenfassung von Texten gibt“, so Mertens Erfahrungen. Bisher zöge man bei avameo allgemein verständliche Texte als Grundlage für die Analysen heran, und weniger Experteninformationen.

Konkrete Erfahrungen aus dem bibliothekarischen Umfeld stellte auch Manfred Hauer von AGI-Information Management Consultants vor. Als Soziologe hinterfragte er zunächst kritisch, ob KI generell ein „Traktor“ des Dienstleistungsgewerbes sei. Betrachte man den strukturellen Wandel der Wirtschaft, könne man hieran zweifeln. Landwirtschaft und Industrie seien im Rückgang begriffen, dagegen hätten Dienstleistungen „am Schreibtisch“ zugenommen. Zu den intellektuellen Dienstleistungen gäbe es bereits Prognosen über automatisierbare Tätigkeiten. Im Bereich der Wissensarbeiter könne KI bald menschliche Arbeit ersetzen, womit trotz moralischer Bedenken zu rechnen sei. Manfred Hauer befürchtet, Information Professionals würden zunehmend zu Solo-Selbständigen mit Direktvermarktung und mobilen Arbeitsgeräten, die trotz Fachkräftemangels wie andere Wissensarbeiter auch „mehr und mehr durch KI substituiert werden“. Politisch gäbe es derzeit kein Konzept, wie

<sup>1</sup> Vgl. hierzu die Nachlese von Andreas Oskar Kempf zu den DGI-Praxistagen <https://www.zbw-mediatalk.eu/de/2018/11/dgi-day-practicum-how-does-artificial-intelligence-change-the-world-of-information-professionals/>

mit dieser Entwicklung nachhaltig umgegangen werden soll. Im zweiten Teil seiner Präsentation ging Hauer auf das Projekt dandelon.com aus seinem Hause ein, welches maschinell Inhaltsverzeichnisse in 200 Sprachen umwandelt und dies als Basis zur Inhaltsererschließung anbietet. „Verschiedene Technologien werden wie beim Lego zusammengebaut“, erläuterte Hauer.

Dass der Mangel an Personalressourcen zur Nutzung von KI zwingt, davon ist Dr. Hidir Aras vom FIZ-Karlsruhe überzeugt. Aber auch er sieht bei seinen Projekten, dass der Information Professional nicht vollständig ersetzt werden könne. Aras referierte über Mining-Verfahren für die Analyse von Patenten. Die Menge an Patenten und die Geschwindigkeit der Anmeldungen nähmen so stark zu, so Aras, dass maschinelle Verfahren benötigt würden, um der Flut Herr zu werden. Er erklärte, wichtige Informationen bei Patentanmeldungen steckten in Texten, z.B. in chemischen Formeln oder Codes. Daraus lassen sich auch Trendanalysen ableiten. Aras arbeitet mit der Open-Source-Software KNIME, deren Analytics Plattform beispielsweise Tagclouds visualisiert. „Semantische Sichtweisen ermöglichen einen neuen Zugang zu Inhalten“, sagte Aras. So könne wertvolles Wissen auch außerhalb des Kontexts von Information Professionals sichtbar und mit weiteren Wissensbasen verknüpft werden. „Zum Glück gibt es kein System, was alles kann, sondern es müssen im Bedarfsfall verschiedene Komponenten orientiert am Ergebnis kombiniert werden“, so seine Erkenntnis. Er bestätigte damit die Erfahrungen von Manfred Hauer und Andreas Mertens.

### KI sieht, was Du nicht siehst

Prof. Dr. Maximilian Eibl von der Universität Chemnitz nutzt KI zur Erschließung von Videos. Die Universität hat einen großen Bestand von Fernsehsendungen aus der Zeit, als es in Sachsen noch 40 Sender gab, die heute nicht mehr existieren. Die Sendungen sind auf VHS-Kassetten gespeichert, die leider für die Archivierung auf Dauer ungeeignet in Kartons gestapelt sind. Dennoch seien die Sendungen mit lokalen Inhalten und Werbungen ein Stück Mediengeschichte, die es Wert sei, aufgearbeitet zu werden. KI sollte helfen, beim Digitalisieren Korrekturen zu ermöglichen. Sprache, Bildinhalte, Texte in Bauchbinden sollten für die Metadatengewinnung OCR-gewandelt werden. Von den Projektergebnissen der Universität könnten auch Medienarchive profitieren.

Die Audioanalyse stand aufgrund des lokalen sächsischen Dialektes auf den Aufnahmen vor großen Herausforderungen, berichtete Eibl humorvoll. Also brauchte es spezielle Trainingsdaten. Besonderes Au-

genmerk richtete das Projektteam auf die Systematisierung von Störquellen bei VHS-Kassetten. Da die Technik und das Wissen aus den 80er Jahren stammen, gab es bisher kein homogenes Regelwerk für Störungen. KI fand hier sieben Phänomene, darunter Textauschen, Flimmern, Verzerrungen. Eibl und seine Studierenden sind auf der Suche nach einem Archiv, welches die VHS-Kassetten übernimmt.

### Die Zeit ist gekommen

Die Veranstaltung der DGI war mit über 70 Teilnehmenden gut besucht und warf auch für Nicht-KI-Spezialisten in nachvollziehbarer Weise Schlaglichter auf konkrete Projekte. Es geht nicht mehr ohne KI, aber auch nicht ohne menschliche Intelligenz. Das konnte man als Erkenntnis aus den Vorträgen mitnehmen. Die Veranstaltung machte deutlicher, wo die Chancen und Grenzen von KI liegen und an welchen Stellen noch menschliche Intelligenz und die Arbeit von Information Professionals vonnöten ist, nämlich wenn als „Gatekeeper“ Ergebnisse in einen Qualitäts- und Entscheidungskontext eingeordnet werden. Information Professionals können sich viel Arbeit von Algorithmen abnehmen lassen. Sie können Erfahrungen mit der Technologie machen und letztlich für Transparenz und Verbesserung der Verfahren sorgen. Im professionellen Anwendungskontext ohnehin eine wichtige Vertrauensfrage.

Der Konferenz am 9. November waren am Vortag öffentliche Sitzungen zweier Arbeitskreise vorausgegangen. Der AK „Bildung und Informationskompetenz“ näherte sich der Frage an, was KI für die Bildung leisten könnte, wie der Zugriff auf offene Bildungsressourcen das Retrieval beeinflusst und wie sich KI mit Informationskompetenz verträgt. Der Arbeitskreis „Terminologie und Sprachfragen“ informierte Interessierte grundsätzlich über die Arbeit des AKs und erarbeitete konkrete Ideen für Kooperationen, wie Terminologie im neuen Informationsumfeld nützen kann. Ausführlichere Ausarbeitungen der Vorträge erscheinen in den nächsten Ausgaben der Zeitschrift IWP – Information – Wissenschaft & Praxis. **I**

Link zu Programm und Präsentationen auf den Webseiten der DGI:

<https://dgi-info.de/dgi-praxistage-2018-programm-der-algorithmus-wo-man-mit-muss/>



**Elgin Helen Jakisch**

U&B Interim-Services, Berlin  
jakisch@ub-interim.de