

KOLUMNE

Im Schatten der Algorithmen

Forschungsdatenmanagement als Fundament intelligenter Systeme

Annette Strauch-Davey



Bild: pixabay.com

► Künstliche Intelligenz (KI) steht im Kontext der Automatisierung menschlicher Intelligenz in Maschinen. Diese sind so programmiert, dass sie wie Menschen denken und lernen und Aufgaben erledigen können, die sonst menschliche Intelligenz erfordern. KI verändert auch das Forschungsdatenmanagement, und das ist auch seit einiger Zeit schon bei mir, in meiner täglichen Arbeit, im Forschungsdatenmanagement angekommen. In den Beratungen bei der Kontakt- und Servicestelle FDM häuften sich die allgemeinen Fragen zu KI in der Forschung: „Was dürfen wir als Forscher, was dürfen wir nicht?“ Wir befinden uns in einer neuen Zeit des Forschungsdatenmanagement. Die „Big Data“-Landschaft¹ ist heute eine andere als noch vor einigen Jahren, weil oft über die natürliche Sprache mit Forschungsdaten interagiert wird.

„Die Integration von KI in Forschungsmethoden verspricht eine neue Ära wissenschaftlicher Entdeckungen, definiert traditionelle Paradigmen neu und treibt den wissenschaftlichen Fortschritt voran.“ KI-gestützte Tools ermöglichen beispielsweise eine automatisierte Erfassung von Forschungsdaten.²

KI als Schlüsseltechnologie ist wichtig für die Wettbewerbsfähigkeit des Wissenschaftssystems in Deutschland.³ KI bietet beim Umgang mit Forschungsdaten Risiken, aber vor allem auch gute Chancen. Herausforderungen sind Fehlinformationen bis hin zu Bias (Verzerrungen) der Forschungsergebnisse, weil die Transparenz oft nicht ganz klar ist. Trainingsdaten bevorzugen bestimmte Gruppen, so zum Beispiel in der Gesundheitsforschung Männer und nicht Frauen, und das kann eben ganz gravierende Folgen haben. In der Gesundheitsforschung handelt es sich dabei oft um die medizinische Versorgung von Frauen. In Studien sind Frauen oft unterrepräsentiert. KI bedeutet für Diagnostik, Therapie und Versorgung ein großes und gutes Potential, aber es müssen bei der automatisierten Auswertung von Daten geschlechtsspezifische Aspekte gewährleistet sein, sonst droht eine Fehlversorgung. So sind Diagnosen von Herzinfarkten oft verzerrt, weil die Systeme, die auf KI basieren, Frauen benachteiligen. Large Language Models (LLMs) sind eine Form von KI, die auf riesigen Textmengen trainiert wurden. Sie verstehen menschliche Sprache und erzeugen Sprache. Bekannte Beispiele sind generative vortrainierte Transformer (generative pre-trained transformers), GPT.

1 Herb, Ulrich. „Zwangsehen und Bastarde: Wohin steuert Big Data die Wissenschaft?“ Information – Wissenschaft & Praxis, vol. 69, no. 2-3, 2018, pp. 81-88. <https://doi.org/10.1515/iwp-2018-0021>

2 KI und FDM - Forschungsdatenmanagement

3 Wissenschaftsministerkonferenz (2025a): Künstliche Intelligenz (KI): Schlüsseltechnologie für Fortschritt und Wettbewerbsfähigkeit des Hochschul- und Wissenschaftssystems. Positionspapier, https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2025/2025_01_31-KIPositionspapier.pdf

„Large Language Models beherrschen das Verarbeiten von komplexen Texten, Fragen und Anweisungen und sind in der Lage, schlüssige Texte mit korrekter Grammatik und Rechtschreibung zu generieren.“, so heißt es bei Fraunhofer.⁴ Texte und Code müssen aber weiterhin von Menschen geprüft werden. Wir dürfen die Kontrolle nicht verlieren.⁵ Modelle programmieren gerne in Python. Large Language Models gehen gut mit menschlichen Sprachstrukturen um. Forschungsinnovationssysteme werden nun durch KI verändert. Forschungsdatenmanagement findet durch Automatisierungen Unterstützung, die aber eine FDM-Kontakt- und Beratungsstelle niemals ersetzen kann. Kollaborative Prozesse können beschleunigt werden, wenn ich auf KI-Tools rechtskonform hinweise. Bisher hat die Nationale Forschungsdateninfrastruktur (NFDI) noch kein Positionspapier, d.h. wie sie beabsichtigt, mit KI und FDM auf nationaler Ebene umzugehen.

mäß der guten wissenschaftlichen Praxis forschen. FAIRe Forschungsdaten, vor allem auch mit KI, sind beim Forschungsdatenmanagement das „A & O“. Besonders das NFDI-Konsortium NFDI4DataScience unterstützt also eine community-getriebene Forschungsdateninfrastruktur für KI. Insgesamt versteht sich die NFDI als ein KI-Enabler, neben den High-Performance- Rechenzentren: „Die NFDI fungiert primär als methodischer und sozialer KI-Enabler. Sie wird KI-Dienste bzw. den Zugang zu KI-Diensten realisieren und die Nutzung, Vermittlung und Analyse von Forschungsdaten in Deutschland auf ein neues Niveau heben.“⁶ Mit dem Schreiben „Information für die Wissenschaft“ fördert die DFG den Aufbau von Datenkorpora zum Trainieren von Künstlicher Intelligenz. Gegenstand der Förderung sind Qualitätskriterien, Nachnutzung (Datenbereinigung, Datenannotation, usw.).⁷

Zu den in dieser Kolumne genannten Bereichen führt das

CHATGPT

Bild: pixabay.com

Das Konsortium NFDI4DS beschäftigt sich aber seit einiger Zeit schon mit KI-Methoden (datengetriebenen und Deep-Learning-Ansätzen). Link zum Konsortium: <https://www.nfdi4datascience.de/>, Link zu Literatur (bzw. Outcome): <https://www.nfdi4datascience.de/outcome/>

Wie Künstliche Intelligenz als Treiber für innovative Forschungsinnovationssysteme genutzt werden kann, wurde Anfang 2025 im folgenden Artikel näher erläutert, auf den ich hier nicht näher eingehen kann, aber den ich empfehlen möchte:

Azeroual, Otmane, Störl, Uta, Rothfritz, Laura, Schöpfel, Joachim, Herb, Ulrich and Borges, Georg. „Datenmanagement 4.0: Künstliche Intelligenz als Treiber für innovative Forschungsinformationssysteme“ *Information – Wissenschaft & Praxis*, vol. 76, no. 1, 2025, pp. 32-42. <https://doi.org/10.1515/iwp-2024-2048>

Cyber Valley ist seit 2016 das größte Forschungszentrum für Künstliche Intelligenz und moderne Robotik in Europa. Hier werden Wissenschaftler gefördert, die ge-

österreichische Cluster Forschungsdaten Ende November 2025 die Veranstaltung **"FAIR & AI: How research data and AI can empower each other"**⁸ durch. Die zentralen Fragen sind auch hier: „Wie können FAIRPrinzipien KI-Anwendungen unterstützen, und wie kann KI-RDMProzesse verbessern?“ Infrastruktur bleibt eine Herausforderung und wie die Interdisziplinarität auf globaler Ebene (European Open Science Cloud/EOSC und mehr) gelingen kann. Der Umgang mit KI ist ein verantwortungsvoller. Er steht an erster Stelle. **I**



Annette Strauch-Davey

Annette Strauch-Davey, M.A. baut aktuell das Forschungsdatenmanagement an der Universität Witten/Herdecke auf. Sie ist dort der Fakultät für Gesundheitsforschung zugehörig.

Annette.Strauch-Davey@uni-wh.de

4 Was sind Large Language Models? Und was ist bei der Nutzung von KI-Sprachmodellen zu beachten? - Blog des Fraunhofer IESE (Zugriff am 16.10.2026)

5 Roberto V. Zicari (Oktober 2025): "Beyond the AI Hype: Guido van Rossum on Python's Philosophy, Simplicity, and the Future of Programming" – *Beyond the AI Hype: Guido van Rossum on Python's Philosophy, Simplicity, and the Future of Programming*. | ODBMS Industry Watch (Zugriff am 16.10.2025)

6 Wissenschaftsrat (2025): Strukturevaluation der Nationalen Forschungsdateninfrastruktur (NFDI); Köln. <https://doi.org/10.57674/wcdc-6d36> (Zugriff am 16.10.2026)

7 DFG - Deutsche Forschungsgemeinschaft - Datenkorpora für Künstliche Intelligenz

8 FAIR&AI: How research data and AI can empower each other - Cluster Forschungsdaten