

Bericht zur Session „Alles ChatGPT oder was? – Chancen und Herausforderungen Künstlicher Intelligenz (KI) in der Evaluation“ auf der 26. Jahrestagung der DeGEval, 14.09.2023

Die Entwicklung von Künstlicher Intelligenz (KI) hat das Potential, den Berufsstand der Evaluation stark zu verändern. In der Evaluationsforschung ist das Thema aber gerade erst im Begriff, größere und regelmäßige Aufmerksamkeit zu erlangen – unter anderem ausgelöst durch den Hype um Large Language Models wie den Text Roboter Chat GPT. In einer **gemeinsamen Session „Alles ChatGPT oder was? – Chancen und Herausforderungen Künstlicher Intelligenz (KI) in der Evaluation“** setzten sich die beiden **DeGEval-Arbeitskreise „Methoden in der Evaluation“** und **„Professionalisierung“** mit den Chancen und Herausforderungen des Einsatzes von KI in der Evaluierungsarbeit und für Evaluierende auseinander.

Mit einem Kurzinput wurde in der von Franziska Heinze¹, Alexander Kocks² (Arbeitskreis Methoden in der Evaluation) und Jessica Prigge³ (Arbeitskreis Professionalisierung in der Evaluation) initiierten und sehr gut besuchten Session (90 Teilnehmende) zunächst ein gemeinsames Verständnis darüber hergestellt, was unter KI zu verstehen ist und reflektiert, wie und unter welchen Bedingungen KI in der Evaluation zur Anwendung kommen kann. In seinem **Impulsvortrag „Künstliche Intelligenz (KI) in der Evaluierungspraxis – von der technologischen Revolution in den Arbeitsalltag“** beleuchtete **Kai Rompczyk**, Deutsches Evaluierungsinstitut für Entwicklungszusammenarbeit (DEval), die jüngsten Fortschritte in dem Bereich KI, insbesondere die Entwicklungen von Sprachmodellen wie ChatGPT. Mit der Erfindung des Prinzips autonom lernender und handelnder KIs, gesteigerter Rechenleistung und Datenverfügbarkeit (Big Data) sowie sinkender technischer Kosten ist die Implementierung von KIs in unterschiedlichsten Anwendungsbereichen möglich geworden und verbreitet sich rasant. In immer mehr Teilbereichen übersteigen die KIs dabei die Kompetenzen von Menschen. Rompczyk demonstrierte die Nutzungsmöglichkeiten von KIs anhand von Beispielen wie automatisierte Aufbereitung von Evidenz, Qualitätssicherung oder Durchführung von statistischen Analysen (z.B. Regressionsanalysen). Dabei machte er deutlich, dass der Einsatz von KIs in allen Evaluierungsphasen nützlich sein kann. Ein großer Mehrwert bestehe darin, dass KIs unzählige Iterationen von analytischen Aufgaben aus unterschiedlichsten Perspektiven durchführen können. Abschließend wies er auf die Risiken der KI in der Evaluierungspraxis hin. Er betonte die wachsende Spannung zwischen Big Data und Datenschutz und fragte, ob die aktuellen Methoden zum Schutz sensibler Daten noch ausreichend sind. Er warnte vor Verzerrungen und Vorurteilen, die durch die Trainingsdaten entstehen können, und der Gefahr der Intransparenz durch die Blackbox-Natur vieler KI-Modelle. Rompczyk sprach auch die Herausforderungen an, die sich aus scheinbar menschlichen Kompetenzen mit nichtmenschlichen Fehlern ergeben, und die potenzielle Erosion des Vertrauens in etablierte Wissensnetzwerke durch den Einsatz von KI-Technologien wie Deepfakes. Er betonte

¹ Deutsches Jugendinstitut, Halle, ak-methoden@degeval.org

² Deutsches Evaluierungsinstitut für Entwicklungszusammenarbeit, Bonn, ak-methoden@degeval.org

³ Christian-Albrechts-Universität zu Kiel, ak-professionalisierung@degeval.org

schließlich die Notwendigkeit bei der Verwendung von KIs in der Evaluierungspraxis die Einhaltung von Evaluierungsstandards zu überprüfen.

Davon ausgehend ermöglichte die Session einen **strukturierten Austausch in Kleingruppen zu Chancen und Risiken des Einsatzes der Technologien** mit Blick auf zentrale Evaluationsstandards (Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit) **sowie zur Frage, was der Einsatz von KI für unsere Professionalität und weitere Professionalisierung als Evaluierende** bedeutet. Die Ergebnisse wurden anschließend im Plenum zusammengetragen:

Zahlreiche **Chancen und Potenziale** von KI wurden genannt. Angesichts voranschreitender technischer Entwicklungen und sich zunehmend verbessernden Leistungen von KI-Lösungen wurden viele Vorteile mit Bezug zu Nützlichkeitsstandards gesehen (z.B. Berichterstattung, umfangreichere Datenanalysen, Rechtzeitigkeit). Weitere Potenziale wurden hinsichtlich der systematischen (Fehler-)Überprüfung und Genauigkeit von Datenauswertungen, bezogen auf Automatisierungs- und damit einhergehende Kostenvorteile oder in Bezug auf Möglichkeiten der systematischen Anonymisierung oder Zugänglichkeit von Evaluationsergebnissen (Fairnessstandards) betont.

Herausforderungen von KI wurden z.T. sehr grundsätzlich diskutiert, ohne dass bereits Lösungen gefunden worden wären. Dies betrifft beispielsweise das Verhältnis von menschlicher und Algorithmen-basierter Erkenntnis oder im Anschluss an den Kurzeintrag die mit der Anwendung von KI einhergehende Intransparenz, die zur Einschränkung des Datenschutzes und der Nachvollziehbarkeit von Ergebnissen führen kann.

Für den Umgang mit den benannten Herausforderungen wurden **erste Lösungen vorgeschlagen**, die sich vor allem auf Weiterbildungs- und Schulungsbedarfe, Erfahrungsaustausch, technische Lösungen (z.B. lokales Hosting) oder die Verständigung auf die Gültigkeit von Evaluationsstandards (Transparenz, Datenschutz, Möglichkeiten und Limitationen der Technologien kennen) bezogen.

Im Hinblick auf die **Evaluation als Profession** wurden Folgen für das Berufsfeld antizipiert, beispielsweise Risiken in der „Wegrationalisierung“ menschlicher Arbeit, sinkende Kompetenzen und Glaubwürdigkeit von Evaluierenden oder eine Stärkung größerer Evaluierungsinstitutionen. Außerdem wurden Bedarfe der Fort- und Weiterbildung in Evaluation formuliert einerseits zur konkreten Anwendung (wie Prompts schreiben), andererseits wird ein reflexiver Umgang mit KI eingefordert (z.B. „Fehlerquellen antizipieren“) und sie dabei auch selbst zum Evaluationsgegenstand gemacht.

Gemeinsam wurde **ausblickend** der Frage nachgegangen, **was es braucht, um Evaluation bzw. Evaluierende im Umgang mit Chancen und Herausforderungen von KI zu stärken**. Vorschläge gingen einerseits in die Richtung, Erfahrungsaustausch zu stärken und zu fördern, eine Ad hoc-Gruppe einzurichten sowie vor allem die (Selbst-)Vergewisserung im Umgang mit Ansprüchen der Evaluationsstandards im Hinblick auf Chancen, Herausforderungen und Professionalisierungsanforderungen zu stärken.