

Dokumentation des digitalen Follow-up-Austauschs zum Thema „KI in der Evaluation: Bedarfe, Austauschformate und Vernetzung in der DeGEval“

Am 28.11.2023 fand initiiert von Franziska Heinze¹, Alexander Kocks² (Arbeitskreis Methoden in der Evaluation) und Jessica Prigge³ (Arbeitskreis Professionalisierung in der Evaluation) im Nachgang einer Session zum Thema Künstliche Intelligenz (KI) in der Evaluation auf der Jahrestagung der DeGEval im September ein erstes, digitales Follow-up-Treffen statt. Mit rund 45 Teilnehmenden aus Deutschland, Österreich, der Schweiz und Italien wurden folgende Fragen in Kleingruppen und gemeinsam in großer Runde diskutiert:

1. Was sind Themen und Fragen in Bezug auf KI in der Evaluation?
2. Welche Formate sind für weitere Vernetzungs- und Austauschaktivitäten in der DeGEval zum Thema KI geeignet?
3. Welche weiterführenden Infos können geteilt werden?
4. Wer kann sich in welcher Form einbringen?

Das große Interesse von über 70 Personen, die sich vorab für die Veranstaltung registrierten, und die rege Teilnahme führten dazu, dass die Veranstalter*innen spontan ein – aus der Perspektive des Datenschutzes durchaus zu kritisierendes – Padlet⁴ zur Unterstützung und Dokumentation nutzten. Dadurch konnte eine Vielzahl an Bedarfen und Themen im Kontext KI, weiterführende Informationen und Ideen für Austauschformate gesammelt werden. Es zeigte sich, dass die Beschäftigung mit KI in einer großen Bandbreite in der Evaluation vorliegt: von denjenigen Evaluatord*innen, die noch keinen Kontakt hatten und neugierig sind z.B. auf Einsatzmöglichkeiten, über einige, die KI für konkrete Aufgaben in Evaluationen einsetzen bis hin zu ganzen Organisationen, die KI systematisch in ihre Strukturen zu integrieren wünschen.

Deutlich wird: Diese Bandbreite an Erfahrungen auch angesichts der dynamischen Entwicklungen rund um KI ‚unter einen Hut‘ zu bekommen, stellt eine Herausforderung dar. Hinweise, wie damit umgegangen werden kann, gaben die Teilnehmenden selbst wie die Sichtung der gesamten Einträge auf dem Padlet, die unten tabellarisch aufbereitet ist, zeigt. Gewünscht wurde sich vor allem ein auf **Erfahrungsberichten basierendes Austauschformat im digitalen Raum**, in dem erfahrungsbasierte Einsatzmöglichkeiten von KI vorgestellt sowie konkrete Problem- und Fragestellungen diskutiert werden können. Die KI-Nutzenden sind dabei explizit aufgefordert, auch offene Fragen, Konflikte oder ‚gescheiterte‘ Anwendungen zu thematisieren. Auch weitere Themen und Formate wurden eingebracht, die in der Tabelle festgehalten sind.

Abschließend diskutierten wir die vorgeschlagenen Themen und Austauschmöglichkeiten. Da sich noch kein aktiver Personenkreis zusammengefunden hat, werden wir als Initiator*innen des Treffens die Aufgabe übernehmen, auf Basis dieser Ergebnisse ein

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

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

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

⁴ Zum Nachlesen: <https://padlet.com/jessicaprigge1/ki-in-der-evaluation-bedarfe-austauschformate-und-vernetzung-4ovcrle7qucyizaz>

weiteres Treffen zu entwerfen, in welches die diskutierten weiterführenden Themen und Austauschformen ebenfalls einfließen werden.

Neustrukturierte Dokumentation der Ergebnisse auf dem Padlet

Themenfelder

Evaluationsethik: Rechtliche Einordnung und Datenschutz

Thema	Likes	Kommentare
Strategien Datenschutz zu sichern	7	
rechtliche Einordnung	3	Risiken für Unternehmen und für Einzel-Gutachter identifizieren und zu Risikomanagement qualifizieren.
<ul style="list-style-type: none"> • Herausforderungen/Risiken der Anwendungen <ul style="list-style-type: none"> ○ Enttabuisierung der Anwendung und gute wissenschaftliche Praxis: wie legt man in Berichten dar, dass KI genutzt wurde und wie diese genutzt wurde ○ Datenschutz (ggf. eigene lokale Lösungen mit ähnlich guten Ergebnissen) ○ was passiert in der "Black Box" 		
Reproduzierbarkeit von Ergebnissen		

Stand der Entwicklung und Anwendungsbeispiele: Was kann bereits wie genutzt werden?

Anwendungsbeispiele		
Konkrete Tools: was kann man bereits nutzen?	7	passendes Tool für den jeweiligen Gegenstand kennen lernen
Systematische Testung von Tools im gesamten Prozess der Evaluierung und Praxisberichte zu den einzelnen Tools	2	
best practice: prompting, prompt hacking	2	Entwicklung „evaluationsspezifischer Prompts“ nach Kategorien (z. B. Phasen einer Evaluation)
<ul style="list-style-type: none"> • Anwendungsmöglichkeiten <ul style="list-style-type: none"> ○ quantitativ ○ qualitativ ○ Berichtswesen 	1	

<ul style="list-style-type: none"> • Datenkompetenz (Befähigung Ergebnisse zu interpretieren)/ Fortbildungsbedarf bzw. Stellenwert in der Ausbildung 		
---	--	--

Weiterführende Fragen und Diskussionspunkte

Blick in die Zukunft: Utopien in der Evaluation		
Nutzung des Begriffes: AI angemessener als KI?		
Strukturierung der Entwicklungen in KI über verschiedene Rollen der Evaluator*innen		
Formulierungshilfen: wie mache ich Einsatz von AI/KI transparent? Wo sind die Grenzen?		
Bsp.: ist Nutzung von deepl schon Nutzung von KI?		
Einsatz von KI auf organisationaler vs. individueller Ebene		
Definition von Standards Wer setzt die Standards? (zunächst definieren: was sind Themen, die hier diskutiert werden sollen? etc., bevor Standards definiert werden) FAQ (fachlich) - auch als Format		
Transparenz und Erkenntnispotential z.B. Konfiguration von KI-Algorithmen und Auswertung von Daten: welche Möglichkeiten und Grenzen gibt es für neue Erkenntnisse, die außerhalb des von uns „erwarteten“ liegen (wie werden Algorithmen „gefüttert“, ggf. auch mit welchen Erwartungen)		

Austauschformate

(Diskussions-)Fragen		
Konkrete Anwendungsbeispiele – Erfahrungsaustausch	6	
Austauschformate möglichst konkret/auf ganz konkrete Fragestellungen beziehen		
knowledge-sharing-Formate	3	„Koordination von Schwerpunkt-AGs“ „praktisches Knowledge Sharing (Welche Probleme werden gelöst, mit welchen Tools, was hat funktioniert, was hat schlecht funktioniert und vor allem: wie sehen KollegInnen Zukunftsentwicklungen)“

		„möglichst auf ganz konkrete Themen/Fragen fokussieren“
Beispiele gemeinsam besprechen und für Probleme gemeinsame Lösungen finden	2	
Austausch zu Problemen auf Abfrage - dafür Pinnwand / schwarzes Brett mit Problemstellungen und Meldemöglichkeiten	2	
Bündelung bestehenden Wissens zu KI für den Evaluationskontext (ohne sich zu verzetteln), Einbeziehung auch anderer Professionen	2	
best practice-Sammlung	4	Im AK Epol haben wir gerade ein Papier "Gütekriterien für Monitoring" übersetzt und als Übersetzer DeepL plus zwei Personen, die es redigiert haben, genannt.
Plattform in der DeGeval um über KI in Austausch zu bleiben	1	
Koordination von AGs zu Schwerpunktthemen durch AK		
Emailverteiler von heute	1	

Weiterführende Formate

Inhalt	Likes	Kommentare
Positionspapier In die Richtung: Jede Anwendung von KI ist transparent und sichtbar zu machen		
Vortragsreihe ... mit Sprechern (ggfs. aus dem (europäischen) Ausland) um neue Impulse zu bekommen	1	
Session auf der nächsten DeGEval-JT <ul style="list-style-type: none"> • Forum, bei dem Erfahrungen ausgetauscht werden können • konkrete Anwendungsbeispiele und Reflexion • Erfahrungen mit prompting 		
Umfragen zur Bedarfserhebung		-

Welche Themen? Welche Personen können / wollen sich einbringen? Oder bei einem einfachen Austauschformat belassen.		
--	--	--

Infomaterial

Inhalt	Kommentare
<p>Einführung</p> <p>Eher inhaltlich, aber: für einen „Grundkurs“ zu AI ist folgender Kurs empfehlenswert: https://course.elementsofai.com/</p>	
<p>Forschung und Entwicklung</p> <p>z. B. Beitrag Jan Hense und Stefan Rädiker in der ZfEv (2023) zur KI in Evaluation</p>	<p>können Sie den Link zum Beitrag zur Verfügung stellen?</p> <p><i>Antwort:</i> Leider kein Open Access Beitrag, kann erworben werden bei Waxmann⁵</p>
<p>Veranstaltungsreihe Forschung</p> <p>Der Raum Tübingen-Heidelberg, das Neckartal ist in Europa führend in der KI-Forschung. Die Grünen Tübingen planen für Anfang 2024 eine Veranstaltungsreihe zum Thema, nicht speziell zu Evaluation. Bei Interesse: https://www.gruene-tuebingen.de/partei/arbeitskreise/ak-kuenstliche-intelligenz-ki</p>	
<p>MERL Tech Working Group</p> <p>Natural Language Processing Working Group kicks off! - MERL Tech</p> <p>internationale Arbeitsgruppe zu NLP in Evaluierungen mit starken Bezügen zu KI</p>	
<p>KI-Gruppe im Netzwerk der Qualitätsmanager*innen österreichischer Universitäten</p>	<p><i>Nachtrag:</i> Es handelt sich um eine geschlossene Gruppe, aber ggf. entstehen hier Initiativen für weiteren Austausch</p>
<p>AEA-Impulse</p> <p>https://aea365.org/blog/artificial-intelligence-ai-using-ai-to-build-evaluation-capacity-by-david-fetterman/</p>	
<p>New Directions for Evaluation: Special Issue: Evaluation and Artificial Intelligence, Volume 2023, Issue 178-179 (Hinweis per Mail von A. von Werthern)</p> <p>https://www.tab-beim-bundestag.de/news-2023-04-studie-zu-chatgpt-fur-den-deutschen-bundestag.php</p>	

⁵ https://www.waxmann.com/waxmann-zeitschriften/waxmann-zeitschriftendetails/?tx_p2waxmann_pi2%5bartikel%5d=ART105501&tx_p2waxmann_pi2%5baction%5d=artikel

(Hinweis per Mail von B. Schober)	
Aktuell wird eine intensive Diskussion zur Nutzung von KI/AI in der qualitativen Forschung und der entsprechenden Methodenlehre im Rahmen der QSF-Liste (Mailingliste Qualitative Sozialforschung) geführt (QSF_L mailing list: QSF_L@lists.fu-berlin.de ; https://lists.fu-berlin.de/listinfo/qsf_l)	