

Prof. Dr. Christian Kellermann

Hochschule für Technik und Wirtschaft Berlin
Deutsches Forschungszentrum für Künstliche Intelligenz
Institut für die Geschichte und Zukunft der Arbeit
christian.kellermann@htw-berlin.de

Deutscher Bundestag

Ausschuss für Digitales

Ausschussdrucksache

20(23)150

19.05.2023

Stellungnahme für den Ausschuss für Digitales im Rahmen der Öffentlichen Anhörung „Generative Künstliche Intelligenz“, 24. Mai 2023

**Thema: Anwendungsmöglichkeiten von Foundation Models und Auswirkungen auf
Arbeit und Beschäftigung**

1. Methodische Grenzen aktueller empirischer Studien

Analog zur Debatte über die Auswirkungen von KI-Methoden auf Arbeit und Beschäftigung vor zehn Jahren sind die veröffentlichten Szenarien zu den Effekten von Foundation Models von hohen Substitutionspotenzialen auf dem Arbeitsmarkt geprägt. Auffällig häufig wird die Zahl von 300 Mio. Jobverlusten genannt, die auf eine Studie der US-Investmentbank Goldman Sachs zurückgeht (Briggs/Kodnani 2023). Wie andere Studien dieser Art auch, werden Berufsdatenbanken (O*NET für die USA, ESCO für die EU) und deren Tätigkeitsbeschreibungen als Datengrundlage für einen Abgleich für die Anwendungsmöglichkeiten von KI herangezogen. Dabei geht es um die Identifikation von Arbeitsaktivitäten, die der KI-Automatisierung ausgesetzt sind. Eine Skala der Datenbank O*NET ordnet Tätigkeiten einem Schwierigkeitsgrad zu. Relevant für die Beurteilung des Schwierigkeitsgrads ist die vorausgesetzte Kompetenz für die Ausführung der Tätigkeit. Diese wiederum variiert erheblich in ihrer Bedeutung je nach Berufskontext. Die Methodik vieler Studien zu Arbeitsmarktszenarien spiegelt dann ein geschätztes Kompetenzniveau der KI mit der Schwierigkeitsskala. In der genannten Studie kann KI Aufgaben bis Level 4 von 7 so erfüllen, wie sie ein Mensch erfüllen würde. Aus diesem Abgleich wird dann das Potenzial an Tätigkeiten ermittelt, die teilweise durch KI automatisiert werden können.

Resarcher von OpenAI haben ebenfalls eine solche Projektion auf der Grundlage ihres LLM angestellt (Eloundou et al. 2023). Sie identifizieren Berufe auf der Grundlage ihrer Übereinstimmung mit LLM-Fähigkeiten. Nach dieser Kalkulation sind ca. 80 Prozent der US-

Arbeitskräfte zu mindestens 10 Prozent ihrer Arbeitsaufgaben von der Einführung von LLMs betroffen. Bei ca. 19 Prozent der Arbeitnehmer:innen sind mindestens 50 Prozent ihrer Aufgaben betroffen. Die „exposure“ betrifft alle Einkommensgruppen, wobei höheres Einkommen und LLM-Anwendbarkeit stärker korrelieren. Wie bei anderen Studien auch, wird von einer Steigerung der Arbeitsproduktivität ausgegangen, bzw. die potenziellen Wachstumseffekte extrapoliert: Je nach Adaptions- und Verbreitungsgrad der Technik wird der Effekt mit bis zu 1,5 Prozent jährliches Produktivitätswachstum über eine Strecke von 10 Jahren modelliert (Briggs/Kodnani 2023).

Auf der Skill-Ebene werden maschinelle Fähigkeiten im Bereich Übersetzung, Klassifizierung, kreatives Schreiben und auch der Code-Generierung als zentral angeführt für das Automatisierungspotenzial. Die Integrierbarkeit von LLM in spezialisierte Anwendungen erweitert ihr Anwendungsspektrum. Komplementäre Technologien (wie der GitHub Copilot für das Codieren) erweitern die Anwendbarkeit über (repetitive) Routineaufgaben hinaus – die Perspektive, die bislang in der ökonomischen Modellierung von Beschäftigungseffekten dominierte. Übersetzer, Schriftstellerinnen, Meinungsforscherinnen oder PR-Fachleute sind nach dem Modell von Eloundou et al. (2023) besonders der generativen KI „ausgesetzt“ (definiert durch eine mindestens 50-prozentige Reduktion der Zeit für die Erledigung einer Aufgabe).

Die Studienlage skizziert ein zunehmendes Maß an komplementärem Technikeinsatzpotenzial in vielfältigen Arbeitskontexten. Eine belastbare Aussage über den Grad der Substitution von Arbeit können die Studien nicht leisten – aus mehreren Gründen: Zum einen erfasst die Datengrundlage auch auf der Ebene von Tätigkeitsbeschreibungen nicht das „stillschweigende Wissen“ (Autor 2015), das es auch bei Routinetätigkeiten gibt, geschweige denn das „Arbeitsvermögen“ (Pfeiffer/Supan 2015), das durch Erfahrung und flexible, fehleranfällige Arbeitskontexte aufgebaut wird. Zudem kann das Beherrschen einer Tätigkeit wiederum eine Voraussetzung für die Erfüllung einer anderen Tätigkeit sein. Diese Komplexität ist anhand der zur Verfügung stehenden Datenbanken nicht abbildbar. Ein zweiter limitierender Faktor ist die Technik selbst. Zum einen ist es nicht möglich, die künftige Anwendbarkeit generativer KI vorherzusagen. Zum anderen muss die Projektion des Ist-Stands der Technik einem Zielzustand der vollständigen Akzeptanz, Implementierung und somit Nutzung der Technik folgen, was aber nicht nur den betrieblichen, sondern auch den (regionalen oder nationalen) regulatorischen Kontext ausblendet.

2. Qualitative Grenzen der Anwendbarkeit

Nicht nur in Deutschland, auch global bleibt die Einführung von KI-basierten Anwendungen hinter den Erwartungen zurück und die Selbsteinschätzung von Unternehmen hinsichtlich weiterer Automatisierung werden zurückgestuft. Die qualitativen Veränderungen von Arbeit aus Unternehmensperspektive zeichnen ein Bild von neuen Tätigkeiten und Arbeitskräftenachfrage nach KI-Spezialistinnen, und teil- oder vollautomatisierten Tätigkeiten wie Kassierer oder Bankkaufleute. Der größte Beschäftigungsaufwuchs wird im Bildungssektor erwartet. Analytisches Denken und kreatives Denken bleiben weiterhin die wichtigsten Fähigkeiten für Arbeitnehmer:innen (WEF 2023).

Am Beispiel der Automatisierbarkeit des Coding lassen sich Potenziale und Risiken der generativen KI gut demonstrieren: Für typische Wettbewerbs-Programmierprobleme erreichen heutige Sprachmodelle häufig Lösungsraten, die auf dem Niveau von oder sogar jenseits von Menschen sind. Perry et al. (2022) haben jedoch für das Sprachmodell unter GithubCopilot (OpenAI davinci-codex) herausgefunden, dass Menschen mit Hilfe des Sprachmodells mehr Sicherheitslücken in ihren Code bauen – aber gleichzeitig glauben, dass sie weniger Sicherheitslücken produzieren. Borji (2023) sammelt eine Reihe von „failures“ von ChatGPT, die den professionellen Einsatz heute noch stark einschränken. Ferner existieren Probleme für Spezialanwendungen, da es in der Regel keine Trainingsdaten im Internet gibt. Daraus ergibt sich die Herausforderung, wie man den Kontext einer Spezialanwendung in ein KI-System bekommt, ohne das System mit vielen Daten neu trainieren zu müssen. Die Anwender im Spezialanwendungsbereich sind in Deutschland vor allem KMU, die entsprechende Kontextinformationen für ein Programm bereitstellen müssten. Diese Informationen sind jedoch sensibel. Sofern der User in größeren Prompts arbeiten muss („ich nutze System X, löse Problem Y“), liefert er sein Problem offen dem jeweiligen Systembetreiber aus.

Das Beispiel zeigt, dass das Anwendungsspektrum nach wie vor beschränkt ist und eine entsprechende Erweiterung mit einer datensicheren Infrastruktur und einem anderen Foundation Modell einhergehen müsste, das datensparsam, kuratiert, multilingual und ressourcensparsam ist (vgl. die Initiativen Large European AI Models und OpenGPTx)¹.

¹ https://leam.ai/wp-content/uploads/2023/01/LEAM-MBS_KIBV_webversion_mitAnhang_V2_2023.pdf; <https://opengpt-x.de>

Ferner wird deutlich, dass generative KI wertvolle Assistenzfunktionen für komplexe Aufgaben übernehmen kann, sofern die Modelle darauf trainiert werden, den Anwender nicht möglichst schnell mit einer Antwort zufrieden zu stellen, sondern im Gegenteil Ambiguitäten und Grenzfälle in der Anfrage des Users aufdecken, um z.B. Sicherheitslücken vorzubeugen („human in the loop“). Schließlich müssen Modelle stärker dagegen abgesichert werden, lizenzgeschützten Inhalt zu kopieren oder private Inhalte aus den Trainingsdaten zu exponieren.

3. Fazit: Qualifikationsoffensive für die Produktivkraft KI

In der Summe geben die neueren Modellierungen der Effekte auf Arbeit wenig Aufschluss über zu erwartende Beschäftigungseffekte, weil sie den Arbeitskontext nicht hinreichend erfassen (können). Die Schätzungen der Produktivitätseffekte wiederum sind wichtig, weil sie das Komplementaritätspotenzial der generativen KI einordnen helfen. Hinter dem Indikator steht die qualitative Veränderung von Arbeit durch den vermehrten oder verbesserten KI-Einsatz, welche in der Folge empirisch untersucht werden müsste – im Sinne einer umfassenden Technikfolgenabschätzung, die insbesondere die Beschäftigten und die Mitbestimmungsakteure mit in den Blick nimmt (Albrecht/Kellermann 2020).

Wie der jüngste Arbeitswelt-Bericht festgestellt hat, geht es bei der Digitalisierung allgemein und der generativen KI im Besonderen um die Erfassung von Kompetenzanforderungen am bestehenden Arbeitsplatz oder innerhalb eines Berufsfeldes, deren Ergänzung durch KI in vielen Fällen mit komplexeren kognitiven Anforderungen einhergeht. Der Rat schreibt zurecht, dass die Anpassung der Beschäftigten an die neuen Kompetenzanforderungen stark vom betrieblichen Kontext abhängt. Im Up-skilling und der individualisierbaren Assistenzfunktion der KI liegt das wesentliche Potenzial einer Humanisierung der Arbeit. Umgekehrt besteht das Risiko, dass Qualifikationen von Beschäftigten in Betrieben ohne entsprechende Weiterbildungs- und Qualifizierungsmöglichkeiten überdurchschnittlich schnell entwertet werden (Rat der Arbeitswelt 2023).

Nicht zuletzt vor diesem Hintergrund ist es notwendig, die Risikobeurteilung im AI Act für eine differenzierte Anwendungsbetrachtung von KI zu ermöglichen. Für den Bildungsbereich würde das beispielsweise bedeuten, dass der Einsatz generativer KI als Assistent für die Lehrinhalterstellung mit einem geringen Risiko zu bewerten wäre (die Verantwortung durch die Nutzerin vorausgesetzt), während eine Beratungsleistung durch Chatbots kritischer zu

betrachten wäre, wie auch eine automatische Notengebung. Für die berufliche Weiterbildung kann die KI beim Matching mit den richtigen Lehrinhalten und -formaten besonders wertvoll sein.

Quellen

- Albrecht, T. / Kellermann, C. (2020). Künstliche Intelligenz und die Zukunft der digitalen Arbeitsgesellschaft, Forschungsförderung der Hans-Böckler-Stiftung, Working Paper 200, Düsseldorf. https://www.boeckler.de/de/faust-detail.htm?sync_id=9132 (zuletzt abgerufen am 18.05.2023)
- Autor, D. (2015). Why are There Still So Many Jobs? The History and Future of Workplace Automation. In: Journal of Economic Perspectives, Nr. 29(3), S. 3–30. <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.29.3.3> (zuletzt abgerufen am 18.05.2023).
- Borji, A. (2023). A Categorical Archive of ChatGPT Failure, <https://arxiv.org/pdf/2302.03494.pdf> (zuletzt abgerufen am 18.05.2023)
- Briggs, J. / Kodnani, D. (2023). The Potentially Large Effects of Artificial Intelligence on Economic Growth, Goldman Sachs Economics Research, March. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf (zuletzt abgerufen am 18.05.2023)
- Perry, N. / Srivastava, M. / Kumar, D. / Boneh, D. (2022). Do Users Write More Insecure Code with AI Assistants? <https://arxiv.org/pdf/2211.03622.pdf> (zuletzt abgerufen am 18.05.2023)
- Pfeiffer, S./Suphan, A. (2015). Der AV-Index. Lebendiges Arbeitsvermögen und Erfahrung als Ressourcen auf dem Weg zu Industrie 4.0. Working Paper 2015 #1 <https://www.sabine-pfeiffer.de/files/downloads/2015-Pfeiffer-Suphan-final.pdf> (zuletzt abgerufen am 18.05.2023)
- Rat der Arbeitswelt (2023). Transformation in bewegten Zeiten. Nachhaltige Arbeit als wichtigste Ressource. https://www.arbeitsweltportal.de/fileadmin/user_upload/awb_2023/Arbeitsweltbericht_2023.pdf (zuletzt abgerufen am 18.05.2023)
- Tyna Eloundou, T. / Manning S. / Mishkin, P. / Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, <https://arxiv.org/pdf/2303.10130.pdf> (zuletzt abgerufen am 18.05.2023)
- WEF, World Economic Forum (2023). Future of Jobs Report, https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf (zuletzt abgerufen am 18.05.2023)