Sprache und KI

SuW-Tagung 2025

– Programm –

Inhalt

Mittwoch – 17.09.2025 – Verstehen, Vertrauen, Verwendung

14 Uhr – Prof. Dr. Olaf Kramer (Tübingen): *Generative KI und Autorschaft*

Generative Künstliche Intelligenz verändert in grundlegender Weise das Konzept von Autorschaft, wie es sich seit dem 18. Jahrhundert herausgebildet hat. Während Autorschaft in der Moderne häufig mit Individualität, Kreativität und Authentizität verknüpft wird, treten wir heute in ein Zeitalter hybrider Autorschaft ein. In diesem neuen Kontext wird die eindeutige Zuschreibung von Intentionalität zunehmend problematisch. KI-generierte Texte werfen Fragen danach auf, wer eigentlich als Autor oder Autorin gelten kann – und in welchem Maße Verantwortung für Inhalte übernommen werden kann und muss.

In meinem Vortrag werde ich über neue Konzepte von Autorschaft im Angesicht generativer KI nachdenken. Ich zeige, wie durch technisch realisierte Textualitätskriterien die Fiktion von Autorschaft erzeugt wird und wie diese fiktionale Zuschreibung dennoch Vertrauen schaffen kann. Durch zielgerichtetes Prompting und neue Formen der Textbearbeitung lässt sich, wie ich zeigen werde, ein gewisser Grad an Kontrolle, Verantwortung und Autonomie zurückgewinnen. Damit verbunden ist auch die Frage nach der Autorität von Texten und den dahinterstehenden Instanzen.

Ein Rückblick auf rhetorische Konzepte von Autorschaft in Antike und Vormoderne zeigt zudem, dass es historisch gesehen durchaus Praktiken gab, bei denen nach festen Regeln vorgefertigte topische Argumentationsmuster genutzt wurden, um Texte zu generieren, man Autorschaft jenseits von Individualität und Authentizität konzipieren kann. Diese historischen Beispiele können heute wichtige Perspektiven für den Umgang mit generativer KI öffnen.

15 Uhr – Prof. Dr. Derya Gür-Şeker (Duisburg-Essen): Künstliche Intelligenz im Spiegel der Medien – Konstruktion von Vertrauen und Skepsis

Der Vortrag „Künstliche Intelligenz im Spiegel der Medien – Konstruktion von Vertrauen und Skepsis“ untersucht die mediale Darstellung von Künstlicher Intelligenz (KI) im Zeitraum vor und nach der Veröffentlichung von ChatGPT 3.5 bis ins Jahr 2025. Dabei werden sowohl klassische Medienberichte als auch Diskurse in sozialen Medien vergleichend analysiert. Ein besonderer Fokus liegt dabei auch auf der Auswertung von Userkommentaren, um Einblicke in öffentlich geteilte Einstellungen zu gewinnen und zu verstehen, wie Vertrauen und Skepsis gegenüber KI medial verhandelt und konstruiert werden. Ziel des Vortrages ist es, zentrale narrative Muster in öffentlichen Diskursen herauszuarbeiten, die Vertrauen und Skepsis gegenüber KI formen.

– 16 Uhr: Kaffeepause –

16:30 Uhr – Prof. Dr. Gerhard Schreiber (Hamburg): „Verstehst Du auch, was Du generierst?“ – KI und die Frage des Verstehens

Generative KI-Systeme wie ChatGPT erzeugen überzeugend menschenähnliche Texte und finden inzwischen breite Anwendung auch in Bildung und Wissenschaft. Doch verstehen diese Modelle auch, was sie generieren und können sie das jemals? Der Vortrag skizziert die rasante Verbreitung und Funktionsweise von KI-Textgeneratoren – von Weizenbaums ELIZA bis zu modernen Large Language Models (LLMs) – und beleuchtet, wie diese Systeme bereits heute unsere Kommunikation und Denkprozesse beeinflussen. Im Mittelpunkt steht die in der aktuellen KI-Forschung geführte Debatte, ob generative KI jemals zu echtem Sprachverstehen fähig sein wird oder letztlich auf eine Simulation menschlichen Verstehens beschränkt bleibt.

17:30 Uhr – Prof. Dr. Friedemann Vogel & Emily Reeh:
„Künstliche Intelligenz“ in der strategischen Kommunikation öffentlicher Diskurse: Der schmale Grat zwischen Diskursfiktion und beobachtbarer Praxis

Neue Medientechnologien sind seit jeher sowohl Gegenstand spekulativer Narrative als auch Teil beobachtbarer Praktiken strategischer Kommunikation. Oft ist der Grat zwischen Diskursfiktion und realen Versuchen kommunikativer Überwachung oder Meinungsmanipulation schmal. Das gilt auch für die Entwicklung und Popularisierung von generativen Sprachmodellen und darauf aufbauenden Applikationen, die mit dem Label „künstliche Intelligenz“ versehen werden. Schon seit fast zehn Jahren finden sich in öffentlichen wie fachlichen Diskursen Behauptungen, automatisierte „Social Bots“ manipulierten öffentliche Debatten und bedrohten die Demokratie. Auch heute finden sich vielerorts „KI“-Topoi, die entweder versprechen, mit automatisierten Text- und Bildgeneratoren Kunden oder Wähler quasi von alleine zu verführen, oder dystopische Warnungen vor Fake News oder „hochintelligenter“ Kriegspropaganda. Dabei zeigen sich werbekommunikative Strategien von KI-entwickelnden Firmen, die hochwertwörtergeprägte Kollokationen wie *Demokratie verteidigen* für die Vermarktung KI-gestützter Waffensysteme nutzen. Hierbei ist von Interesse, inwiefern sich der strategische Sprachgebrauch im Werbediskurs von anderen Teildiskursen unterscheidet: Während z.B. Entwickler autonomer Waffensysteme die Effizienz und Präzision jener „technologischen Revolution“ hervorheben, warnen Akteure aus dem Politik- und Rechtsdiskurs vor Risiken der KI-gestützten Kriegsführung. Gleichzeitig verändert der auch für Laien niedrigschwellige Einsatz generativer Text-, Bild- und Sound-Generatoren die Möglichkeitsbedingungen gegenwärtiger Diskurse: von der adressatenorientierten Textoptimierung über die Erstellung fotorealistischer Schlagbilder bis hin zu Voice- und Facecloning in Kriegspropaganda. Im Anschluss an die Arbeiten der Forschungsgruppe *Diskursmonitor* (https://diskursmonitor.de) gibt der Vortrag einen Einblick in die Narrative und kommunikationsstrategischen Praktiken unter dem Eindruck „künstlicher Intelligenz“ und diskutiert mögliche Folgen für öffentliche Wahrheitsregime.

Donnerstag – 18.09.2025 – Schreiben mit KI und Wissenschaftskommunikation

9 Uhr – Prof. Dr. Jan Georg Schneider (Landau): Referenz und Autorschaft am Beispiel LLM-generierter Aufsatzbewertungen

In diesem Vortrag wird der linguistische und sprachphilosophische Status KI-generierter Texte grundsätzlich diskutiert. Während wir bislang so sozialisiert sind, dass wir bei sprachlichen Gebilden, die *als intelligent lesbar* sind, automatisch einen intelligenten Autor bzw. eine intelligente Autorin ‚dahinter‘ voraussetzen, können wir diese enge Verbindung im Zeitalter der LLMs in zunehmendem Maße nicht mehr einfach unterstellen (Durt/Froese/Fuchs 2023). In diesem Sinne spreche ich von "intelligiblen Texturen" (Schneider 2024). Die als intelligent lesbaren Gebilde sind als *Produkte* kaum noch oder gar nicht mehr von menschengemachten, autorisierten Texten unterscheidbar, die Lern- und Gebrauchs*prozesse* unterscheiden sich jedoch grundlegend – besonders im Hinblick auf Referenzakte (Austin 1975, Elgin 1983). Welche Konsequenzen hat dies für unser Verstehen und unsere generelle Auffassung von Geschriebenem sowie für unsere Vorstellung von Autorschaft und damit zusammenhängende Fragen der Verantwortung für verbale Produkte? Diese grundsätzliche Frage wird im Vortrag am Beispiel KI-generierter Aufsatzbewertung diskutiert (Schneider/Zweig 2022, Schneider 2024). Obwohl hier natürlich auch viele andere Textsorten möglich wären, eignet sich die Aufsatzbewertung besonders gut, da hier in hohem Maße Urteilkraft sowie die Bezugnahme auf andere Texte, nämlich die zu beurteilenden, und auch auf deren Wahrheitsgehalt, benötigt wird.

Literatur
Austin, John L. 1975. How To Do Things With Words. 2. Aufl. Oxford: Oxford University Press [1. Aufl. 1962].
Durt, Christoph, Tom Froese & Thomas Fuchs. 2023. Large Language Models and the Patterns of Human Language Use: An Alternative View of the Relation of AI to Understanding and Sentience [PhilSci Archive Preprint] https://philsci-archive.pitt.edu/22744/
Elgin, C.Z. (1983). With Reference to Reference. Indianapolis: Hackett Publishing Co, Inc.
Schneider, Jan Georg (2024): Intelligible Texturen. Welche Rolle kann ChatGPT bei der Aufsatzbewertung spielen? In: VK:KIWA. https://zenodo.org/records/10877034 doi: 10.5281/zenodo.10849262
Schneider, Jan Georg & Katharina A. Zweig. 2022. Ohne Sinn. Zu Anspruch und Wirklichkeit automatisierter Aufsatzbewertung. In: Sarah Brommer, Kersten Sven Roth & Jürgen Spitzmüller (Hgg.): Brückenschläge: Linguistik an den Schnittstellen. Tübingen: Narr Francke Attempto (Tubinger Beitrage zur Linguistik, 583), 271–294.

10 Uhr – Dr. Daniel Knuchel (Zürich/Genf): (Post-)*Digitales Schreiben als Assemblage? Sozio- und diskurslinguistische Perspektiven auf KI-basierte Schreibpraktiken im Bildungskontext*

In aktuellen Debatten rund um KI-basiertes Schreiben werden normative Vorstellungen darüber sichtbar, was Schreiben ist, sein soll – und sein darf. Der Vortrag untersucht entsprechende Schreibideologien anhand zweier Datenquellen: Interaktionen von Schüler:innen, die generative KI bei einer Schreibaufgabe nutzen, sowie institutioneller Leitfäden, die den Einsatz von KI reglementieren. Schreiben wird dabei praxistheoretisch als soziale Praktik verstanden. Im Zuge der Auseinandersetzung mit dem Material zeigt sich, dass Schreiben nicht mehr als klar umrissene Handlung eines autonomen Subjekts beschreibbar ist, sondern als relationale, situative Ko-Produktion. Begrifflich wird dies im Anschluss an posthumanistische Theorien als Assemblage gefasst und im Kontext postdigitaler Bildungskulturen reflektiert.

– 11 Uhr: Kaffeepause –

11:30 Uhr – Dr. Maurice Fürstenberg (München): *Warum KI (nicht) versteht – Funktionsprinzipien und Anwendungsfälle künstlicher Intelligenzillusionen*

In Debatten um Künstliche Intelligenz zeigt sich häufig ein Sprachgebrauch, der kognitionspsychologische Konzepte wie „Verständnis“ auf Technologien wie generative Transformernetzwerke überträgt. So wird KI-Systemen unterstellt, natürliche Sprache mitsamt ihrer Konzepte zu verstehen (u. a. Gastaldi 2021, Dhingra et al. 2023). Dieser Gebrauch beruht auf einem anthropomorphisierenden Missverständnis und einer Überschätzung der technischen Funktionsweise, die durch entsprechendes Framing verstärkt wird. Eine zentrale Ursache liegt in einer kognitiven Verzerrung, bekannt als Pareidolie (Liu et al. 2014): der Neigung, in menschenähnlichen Mustern Menschliches zu erkennen. Wenn KI-Systeme menschlich wirkende Sprache erzeugen, erkennen wir oft nicht die Illusion als solche – wir erliegen einer „intendierten Immersion“ (Fürstenberg/Müller 2024), die durch wirtschaftliche Interessen gezielt verstärkt wird. Aus theoretischer Sicht kann KI höchstens eine Illusion von Verständnis erzeugen, da ihr ein echtes Weltmodell fehlt (Ha/Schmidhuber 2018; Müller/Fürstenberg 2023). Dennoch bleibt die Frage relevant, ab wann Illusion und Wirklichkeit verschwimmen – eine Frage, die bereits Platon beschäftigte (Ebert 1974) und die für explainable AI zentral ist (u. a. Arrieta 2020, Speith 2024). Der Vortrag diskutiert diese Frage anhand der Funktionsweise generativer Transformernetzwerke und eines konkreten Anwendungsfalls – KI-generiertes Textfeedback. Ziel ist es nicht, endgültige Antworten zu geben, sondern Anregungen aus verschiedenen Disziplinen (Informatik, Psychologie, Philosophie, Linguistik, Deutschdidaktik) zu integrieren und einen interdisziplinären Forschungseinblick zu ermöglichen.

Literatur:

Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Dhingra, S., Singh, M., Malviya, N., & Gill, S. S. (2023). Mind meets machine: Unravelling gpt-4's cognitive psychology. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(3), 100139.

Ebert, T. (1974). *Meinung und Wissen in der Philosophie Platons: Untersuchungen zum "Charmides", "Menon" und "Staat"*. Berlin, New York: De Gruyter. https://doi-org.emedien.ub.uni-muenchen.de/10.1515/9783110834826

Fürstenberg, M. & Müller, H.-G. (2024). KI im Deutschunterricht. Funktionsprinzipien und kompetenzbezogene Einsatzmodelle. *Der Deutschunterricht 5*, 2–13.

Gastaldi, J. L. (2021). Why can computers understand natural language? the structuralist image of language behind word embeddings. *Philosophy & Technology*, *34*(1), 149–214.

Ha, D., & Schmidhuber, J. (2018). Recurrent world models facilitate policy evolution. *Advances in neural information processing systems*, *31*.

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex*, *53*, 60–77.

Müller, H. G., & Fürstenberg, M. (2023). Der Sprachgebrauchsautomat. Die Funktionsweise von GPT und ihre Folgen für Germanistik und Deutschdidaktik. *Mitteilungen des Deutschen Germanistenverbandes*, *70*(4), 327–345.

Speith, T., Crook, B., Mann, S., Schomäcker, A., & Langer, M. (2024). Conceptualizing understanding in explainable artificial intelligence (XAI): an abilities-based approach. *Ethics and Information Technology*, *26*(2), 40.

12:30 Uhr – Dr. Vasco Schmidt (SAP, Walldorf): *Vom Verschwinden der Oberfläche – linguistische Umbrüche im technischen Schreiben durch genAI*

Mein Vortrag behandelt die disruptiven Veränderungen, die die generative künstliche Intelligenz (genAI) in der Software-Entwicklung mit sich bringt. Dies erfolgt am Beispiel meiner Abteilung aus technischen Redakteur:innen und UX-Designer:innen bei der SAP SE. Auf den ersten Blick bietet genAI Hilfsmittel, um Inhalte zu recherchieren, Texte zu erstellen und zu redigieren. Die von vielen technischen Redakteur:innen so geliebte Arbeit an der Textoberfläche, das Gestalten situations- und adressatenspezifischer Texte, scheint genAI zu übernehmen, wodurch diese Praxis aus dem Alltag der Redakteur:innen zu verschwinden beginnt. Tätigkeiten zu Inhalten und Metadaten sowie Prozessen und Technologien rücken ins Zentrum. Die UX-Designer:innen erleben ebenfalls ein Verschwinden der Oberfläche. Mit genAI scheint sich der Chat als Interaktionsmuster durchzusetzen. Das von Designer:innen so geliebte Entwerfen und Optimieren graphischer Benutzeroberflächen verschwindet zugunsten des Designs einer auf Text basierender User Experience. Ich möchte die Rolle die Linguistik in diesem Wandel ausloten und aufzeigen, wie sie Reflexionswissen und damit Anregungen für die Praxis in der technischen Redaktion und im UX Design liefern könnte – und damit Hilfestellungen für zwei Disziplinen, die sich gemeinsam neu erfinden müssen.

– 13:30 Uhr: Mittagspause –

15 Uhr: Prof. Dr. Annette Leßmöllmann (KIT Karlsruhe): *KI und Common Ground in der Wissenschaftskommunikation*

In der sprachlichen Interaktion zwischen Menschen, aber auch zwischen Menschen und KI spielt das Mit-Verhandeln eines Common ground eine wesentliche Rolle: Annahmen über geteiltes Wissen oder über geteiltes Commitment bezüglich Gesprächsregeln beeinflussen den Verlauf der Interaktion. Dabei weist z.B. Stalnaker der Wahrheit als Orientierungspunkt eine prominente Rolle zu: In seiner Analyse des Common ground werden geteilte Annahmen von den Teilnehmenden als wahr akzeptiert. Andere Ansätze fokussieren darauf, dass Common ground über eine kommunizierte Wertekonsonanz und ein „Einschwingen" (attunement) zwischen den Interaktanten zustande kommt (Beaver/Stanley 2023). Hier spielt Wahrheit der Annahmen keine

wesentliche Rolle mehr. Gleichzeitig könnte dieser Ansatz aber den Erfolg von Kommunikaten erklären, die Wahrheit zwar als zu vernachlässigende Größe behandeln, aber große Wirkung entfalten. Für die Wissens(chafts)kommunikation, etwa auch den Wissenschaftsjournalismus, ist der Wahrheitsbezug konstitutiv. Damit ist die Frage, welche Theorie des Common ground für die Wissenschaftskommunikation wichtig ist, sehr relevant. Sie wird verschärft durch die Nutzung generativer KI sowohl in der Produktion als auch in der Rezeption von Wissenschaftskommunikation, denn Chatbots und andere Tools haben keinen intrinsischen Wahrheitsradar, können aber, allein durch Masse und Gestaltung, stark wirken – als persuasive Agenten, die Faktenbezug suggerieren, ohne diesen einzulösen. Im Vortrag stelle ich die aktuellen Herausforderungen für die Wissenschaftskommunikation vor und versuche mich an Lösungsvorschlägen.

– 16 Uhr: Kaffeepause –

16:30 Uhr – Prof. Dr. Katharina Zweig (Kaiserslautern): *Wie kann das Sprechen über informatische Berechnungen gelingen?*

Wann immer der Computer etwas berechnet, muss dieses Ergebnis kommuniziert werden. Was besagt das auf dem Bildschirm produzierte Zeichen? Ist es mehr als "Dies ist das Ergebnis der Berechnung"? Anhand verschiedener Beispiele legt Katharina Zweig in ihrem Vortrag dar, dass die eher implizite Kommunikation von informatischen Berechnungen zu Missinterpretationen führen kann. In Bezug auf Austins Sprechakttheorie fragt sie, unter welchen Bedingungen die Kommunikation informatischer Berechnungen gelingen kann.

– 17:30 Uhr Pause –

18 Uhr – Podiumsdiskussion: *WissKomm und KI*

- **Dr. Maria Becker (Moderation)**
- **Prof. Dr. Leßmöllmann**
- **Prof. Dr. Katharina Zweig**
- **Gesine Born**
- **Prof. Dr. Derya Gür-Şeker**

<u>Freitag – 19.05.2025 – Konzept, Kommunikation, Interaktion</u>

9 Uhr – Marcel Kückelhaus (Heidelberg) – *Menschliche Künstliche Intelligenz. Konzeptualisierung von KI im Mediendiskurs*

Roboter, Monster, Götter – journalistische Darstellungen von Künstlicher Intelligenz greifen häufig auf mythische und anthropomorphe Bilder zurück. Der mediale Diskurs über KI ist geprägt von Ambivalenz: Während einige Beiträge technikbegeistert oder alarmistisch ausfallen, zeichnen andere ein differenzierteres. Gemeinsam ist ihnen jedoch die Zukunftsorientierung: Was bedeutet es für die Gesellschaft, wenn sich der technologische Fortschritt im KI-Bereich fortsetzt?

Das Entstehen von KI-Mythen lässt sich – in Anlehnung an Hans Blumenberg – verstehen als „kreative Reaktion auf eine zugrundeliegende Bedrohung durch eine den Menschen überfordernde Wirklichkeit“ (Waldow/Forrester 2023: 90). Zwar steht die Technologie im Fokus des Diskurses, doch wird dabei latent die Fragen nach dem „Menschsein“ verhandelt: Was ist der Mensch? Was unterscheidet ihn noch von der Maschine? Welche Rolle nimmt er in einer Gesellschaft ein, in der KI seine Aufgaben übernimmt?

Der Vortrag untersucht, wie die Beziehung zwischen Mensch und Maschine im journalistischen KI-Diskurs konzeptualisiert wird. Anhand einer qualitativen Analyse journalistischer Beiträge wird dargelegt, wie Vorstellungen von Mensch und Maschine sprachlich und bildlich gefasst werden.

Literatur:

Waldow, Stephanie/Forrester, Eva (2023): Mythos und Mythos-Theorie. Formen und Funktionen. Eine Einführung. Berlin: Erich Schmidt Verlag (Grundlagen der Germanistik, 66).

10 Uhr – PD Dr. Nina Kalwa (Karlsruhe): *Künstliche Irritation. Zur Aushandlung von Echtheit in digitaler Kommunikation*

Generative künstliche Intelligenz erzeugt unter anderem (Bewegt-)Bilder, die nicht immer sofort als künstlich erzeugt erkennbar sind und auch nicht sein sollen, etwa wenn Desinformation durch Deepfakes verbreitet wird (vgl. z.B. Pawelec 2022). In der digitalen Kommunikation werden Menschen in den letzten Jahren zunehmend mit Bildern und Videos konfrontiert, die nicht durch das Fotografieren oder Filmen von *echten* Menschen entstanden sind, sondern in denen *echte* Menschen, meist Individuen, mittels künstlicher Intelligenz imitiert werden. Dann steht beispielsweise zur Diskussion, ob wirklich die bekannte Politikerin oder der bekannte Schauspieler zu sehen ist. Allein das Wissen um die Existenz von KI-generierten (Bewegt-)Bildern scheint das Vertrauen in „digitale Kommunikate“ (Dang-Anh 2024: 459) zu erschüttern. Gleichzeitig wird laut Krämer (2021: 36) durch die Digitalisierung „die Anbindung von Wahrheit an Vertrauen immer unauflöslicher“. Der Vortrag widmet sich der Analyse der Diskussionen um (vermeintlich) künstlich erzeugte digitale Kommunikate und erörtert, welche Aspekte solcher Kommunikate die digital Kommunizierenden irritieren, wie die Kommunizierenden diese Irritationen äußern und welche Schlussfolgerungen sich dabei bezogen auf gesellschaftliche *common grounds* (Clark/Brennan) ziehen lassen.

Literatur:

Clark, Herbert H./Brennan, Susan E. (1991): Grounding in communication. In: Resnick L.B., Levine J.M., Teasley S.D. (ed.): Perspectives on Socially Shared Cognition.Washington, DC, 127–149.

Dang-Anh, Mark (2024): Remixpraktiken in digitalen Kommunikaten. In: Androutsopoulos J., Vogel F. (ed.): Handbuch Sprache und digitale Kommunikation (Handbücher Sprachwissen (HSW) 23), 455–475. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110744163-022.

Krämer Sybille (2021): Der Verlust des Vertrauens. Medienphilosophische Perspektiven auf Wahrheit und Zeugenschaft in digitalen Zeiten. In: Schicha C., Stapf I., Sell S. (ed): Medien und Wahrheit. Medienethische Perspektiven auf Desinformation, Lügen und „Fake News". Kommunikations- und Medienethik 15, 1, Nomos Baden-Baden, pp 25–42. doi.org/10.5771/9783748923190

Pawelec, Maria (2022): Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *DISO* **1**, 19 (2022). https://doi.org/10.1007/s44206-022-00010-6

– 11 Uhr: Kaffeepause –

11:30 Uhr – Dr. Milena Belosevic & Prof. Dr. Henrik Buschmeier (Bielefeld): *Vertrauen in der Mensch-Maschine-Interaktion am Beispiel von LLM-basierten Chatbots*

Vertrauen ist die Grundvoraussetzung aller sozialer Beziehungen. Es ist nicht nur ein Grundprinzip zwischenmenschlicher Interaktion, sondern spielt auch eine zentrale Rolle für die Art und Weise wie sich die Mensch-Maschine Interaktion entwickelt. Da die Bedeutung von Vertrauen in der Mensch-Maschine Interaktion mit der zunehmenden Popularität von Großen Sprachmodellen (LLM) zunimmt, nimmt der Vortrag diesen Aspekt von Vertrauen am Beispiel von LLM-basierten Chatbots näher in den Blick.

Genauer gesagt wird ein Modell zur linguistischen Untersuchung von Vertrauen in so-genannte KI-Chatbots vorgestellt und an Fallstudien implementiert. Das Modell basiert auf dem Konzept der Kalibrierung von Vertrauen (Muir, 1994), das in den Studien zur Mensch-Maschine Interaktion Anwendung fand (Metzger et al., 2024), bisher jedoch nicht für (diskurs)linguistische Untersuchungen operationalisiert wurde. Diesem Konzept zufolge, sollen Menschen weder zu viel noch zu wenig Vertrauen in automatisierte Systeme haben, sondern ein gesundes Maß an Misstrauen (Visser et al., 2023) mitbringen.

Ausgehend von der Annahme, dass Vertrauen sowohl ein Grundprinzip der Interaktion als auch ein diskursiv hervorgebrachtes Phänomen ist (Muntigl et al., 2024), widmet sich dieser Vortrag der Frage, ob und wie bzw. mit welchen sprachlichen Mitteln das Vertrauen zwischen Nutzer:innen und KI-Chatbots auf der Ebene der medialen Berichterstattung und der direkten Interaktion konstruiert und gegebenenfalls kalibriert wird. Auf beiden Untersuchungsebenen werden mit qualitativen und quantitativen Methoden linguistische Indikatoren identifiziert die auf ein gesundes Maß an Miss- trauen oder übermäßiges Vertrauen hinweisen. Abschließend wird auch diskutiert, wie ein kritischer Umgang mit blindem Vertrauen in KI Chatbots mit linguistischen Methoden gefördert werden kann und welchen Beitrag der vorgestellte Ansatz für eine kulturwissenschaftlich-mediale Perspektive auf Vertrauen in KI (vgl. Gür- Şeker, 2024) leisten kann.

Literatur

D. Gür-Seker. Vertrauen in KI – kulturwissenschaftlich mediale perspektiven. In *Vertrauen in Künstliche Intelligenz*, pages 241–254. Springer, Wiesbaden, 2024.

L. Metzger, L. Miller, M. Baumann, and J. Kraus. Empowering calibrated (dis-)trust in conversational agents: A user study on the persuasive power of limitation disclaimers vs. authoritative style. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, 2024. ACM. doi: 10.1145/3613904.3642122.

B. M. Muir. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

P. Muntigl, C. Scarvaglieri, J. De Wilde, K. Bührig, and A. Wamprechtshammer. Trust in interaction studies. *Front. Commun.*, 9, 2024. doi: 10.3389/fcomm.2024.1448110.

R. Visser, T. M. Peters, I. Scharlau, and B. Hammer. Trust, distrust, and appropriate reliance in (x)ai: a survey of empirical evaluation of user trust, 2023. URL https://arxiv.org/abs/2312.02034.

12:30 Uhr – Dr. Tamara Bodden (Kassel): *„er /sie also bixby sagt das habe ich leider nicht verstanden". Genus, Sexus und Geschlecht im Sprechen über KI*

Welche Rolle spielt Geschlecht bei KI und wie wird dieses über sprachliche Mittel erzeugt? Bereits seit Jahrzehnten wird die Problematik von stereotypen geschlechtlichen Darstellungen von KI, beispielsweise in Namen oder Interfaces kritisch diskutiert und untersucht. Technische Artefakte

können nämlich dazu beitragen, überholte Geschlechterstereotype zu perpetuieren. Während aktuellere KI wie ChatGPT oder MyAI teils weniger geschlechterstereotyp gestaltet werden als vorangehende Modelle, werden trotzdem insbesondere von Nutzer*innen immer wieder geschlechtsindizierende Personalpronomen oder Kosenamen verwendet. Dies hängt unter anderem mit dem Eliza-Effect, einem Anthropomorphisieren von Technik mithilfe von Metaphern zusammen. Der Vortrag untersucht basierend auf einem Korpus aus englisch- und deutschsprachigen Texten das Sprechen über (mehr oder weniger anthropomorph gestaltete) KI. Es wird davon ausgegangen, dass die Zuschreibung von Geschlecht auf verschiedenen Ebenen, in der Entwicklung, der Vermittlung in den Medien oder erst bei den Nutzern selbst stattfinden kann.