

Anthropomorphisierung – Wenn KI vermenschlicht wird!

Erarbeitet von

Dr. Maike Meyer, Dr. Jacqueline Klusik-Eckert und Marie von Lobenstein M.A.

Das Thema „Anthropomorphisierung“ ist äußerst komplex und wird aus verschiedenen Fachdisziplinen und Perspektiven betrachtet. In diesem Video haben wir daher eine Auswahl getroffen, um einen ersten Überblick zu geben. Eine vollständige Behandlung des Themas würde den Rahmen der Lehrveranstaltung sprengen und ist im Rahmen dieses Kurses nicht möglich.

Lernziele	2
Einstieg	2
AnthropoWAS? Was ist denn eigentlich Anthropomorphismus?	2
AnthropoWarum? Warum anthropomorphisieren wir?	3
Auswirkungen von Anthropomorphismus	3
Das Vermenschlichungs-Problem	4
Anthropomorphism by Design:	5
Anthropomorphe Sprache:	5
Take-Home Message	6
Quellen	6
Disclaimer	8

Lernziele

- Du kannst Anthropomorphismus definieren
- Du kannst erklären, warum Menschen anthropomorphisieren
- Du kannst Auswirkungen von Anthropomorphismus nennen
- Du kannst zwei Ausprägungen von Anthropomorphismus: Anthropomorphism by Design und anthropomorphe Sprache ausführen

Inhalt

Einstieg

Na, mal Hand auf's Herz: Hast du schonmal mit deinem Drucker geschimpft? Mit deinem Fahrrad? Oder vielleicht mit einem Navigationssystem? Hast du schonmal auf Steinen Gesichter entdeckt oder dich bei dem Gedanken erwischt, das Auto schaue grimmig drein? Oder hast du dich vielleicht schon mal bei deinem Staubsaugerroboter entschuldigt, weil du ihn aus Versehen angerempelt hast? Vermutlich ist sowas oder etwas Ähnliches allen von uns schonmal passiert. Also, dass wir Gegenstände vermenschlicht haben, denn das machen wir ziemlich häufig.

Quelle [1]

Der Fachbegriff dafür lautet Anthropomorphismus. Ja, nicht so griffig, dieses Wort, vor allem, wenn man es mehrmals hintereinander sagen muss ...

AnthropoWAS? Was ist denn eigentlich Anthropomorphismus?

Anthropomorphismus leitet sich von den griechischen Wörtern "anthropos" (was „Mensch“ bedeutet) und „morphe“ (was „Gestalt“ oder „Form“ bedeutet) ab.

Quelle [2]

Unter Anthropomorphismus versteht man die Tendenz, menschliche Eigenschaften – wie Charakteristiken, Motivationen, Intentionen oder mentale Zustände einer nicht-menschlichen Entität oder Objekten zuzuschreiben.

Quelle [3]

Bei diesen nicht-menschlichen Entitäten oder Objekten kann es sich um alles mögliche handeln, beispielsweise um Tiere oder Pflanzen, aber auch um Computer, Roboter, Autos und fast allem, was dir noch so einfällt. Hier mal ein paar Beispiele:

Einblendung Wall-e, Clippy, Wilson, Siri etc.

Quelle [4]

AnthropoWarum? Warum anthropomorphisieren wir?

Laut Epley machen wir das aus zwei Gründen.

Erstens: Soziale Motivation – soziale Bindung, sozialer Kontakt und Zustimmung von anderen

Quelle [5]

Wir vermenschlichen beispielsweise unsere Haustiere, weil wir das Bedürfnis haben, eine soziale Verbindung zu anderen aufzubauen und aufrechtzuerhalten. Nichtmenschliche Agenten, also mein Hund zum Beispiel, können dieses Bedürfnis scheinbar mühelos erfüllen. Es fällt uns eventuell sogar leichter diese Verbindung mit Haustieren und nicht mit anderen Menschen aufzubauen. Eine Person, die sich einsam fühlt, anthropomorphisiert, um das Gefühl der Entfremdung zu lindern.

Zweitens: Effektnanzmotivation – die Motivation, effektiv mit seiner Umgebung zu interagieren

Quelle [6]

Wir vermenschlichen, weil wir erfolgreich und ohne Unsicherheiten mit unserer Umwelt umgehen wollen. Epley argumentiert, dass wir anthropomorphisieren, kann uns dabei helfen unsere unsichere Welt etwas verständlicher zu machen. Wir haben das Verlangen, Unsicherheit und Mehrdeutigkeit zu reduzieren, zumindest teilweise mit dem Ziel, ein Gefühl von Vorhersehbarkeit und Kontrolle in unserer Umgebung zu erlangen. Wenn wir z. B. autonome Fahrzeuge anthropomorphisieren, könnte das aus dem Wunsch resultieren, das Verhalten des Fahrzeugs zu verstehen und vorherzusagen, Unsicherheit zu reduzieren und das Gefühl der Kontrolle über die Fahrzeugumgebung zu erhöhen.

Auswirkungen von Anthropomorphismus

Das Vermenschlichen von KI-Systemen oder auch von anderen Maschinen kann unser mentales Modell des jeweiligen Systems beeinflussen, also wie wir das System verstehen und wie wir glauben, dass es funktioniert.

Quelle [7]

Aber auch unsere Wahrnehmung von Systemen kann dadurch beeinflusst werden, also beispielsweise, wie wir KI-Systeme annehmen oder ob wir sie benutzen.

Quelle [8]

Es gibt einige Hinweise darauf, dass wir anthropomorphisierte Maschinen oder Systeme positiver wahrnehmen als nicht-vermenschlichte Systeme, ihnen also beispielsweise mehr Vertrauen schenken oder sie für vertrauenswürdiger halten.

Quelle [9][10][11][12]

Es gibt beispielsweise auch Studien, die nahelegen, dass die Aktionen von autonomen Fahrzeugen im Straßenverkehr anders bewertet werden, wenn das Fahrzeug anthropomorphe Eigenschaften besitzt und wir ihnen beispielsweise weniger Schuld für einen Unfall zuschreiben, als menschlichen Fahrenden.

Quelle [13][14][15]

Es gibt aber durchaus auch Hinweise auf negative oder gegenteilige Effekte der Vermenschlichung von Systemen auf unsere Wahrnehmung.

Das Vermenschlichungs-Problem

Bereits bei den ersten Interaktionsbegegnungen von Menschen mit einem Computerprogramm im Jahr 1966 ist aufgefallen, wie schnell die Wissenschaftler*innen in dem Sprachprogramm ELIZA etwas Menschliches gesehen haben. Der Entwickler Joseph Weizenbaum (MIT) fand diese Beobachtung besorgniserregend und mahnte vor einer Aura des Magischen, die man den Rechenprogrammen austreiben müsse.

Quelle [16]

Diese Mahnung wurde nicht gehört oder bewusst überhört. Mit dem Ziel, die Chatbots in ihrem Kommunikationsstil Menschen immer ähnlicher zu machen, haben wir heute hinter den Anwendungen wie Alexa oder Chat-GPT4 Modelle, die unsere Wahrnehmung an der Nase herumführen. Wir müssen uns daher immer bewusst machen, dass diese Assistenzsysteme und Chatbots nicht fühlen, nicht intelligent im Sinne von menschlicher Intelligenz sind, dass sie nicht denken, nicht verstehen, nicht wahrnehmen und nicht halluzinieren. Das Einzige, was sie können, ist hervorragend rechnen. Sehr, sehr schnell rechnen. Die Ergebnisse empfinden wir dann oft als richtig, meistens nützlich und bisweilen halten wir sie für besonders klug oder empfindsam.

Quelle [17][18]

Wir können hier zwischen zwei Ausprägungen des Anthropomorphismus unterscheiden: zwischen anthropomorpher Sprache und etwas, was im Englischen als „Anthropomorphism by Design“ bekannt ist.

Anthropomorphism by Design:

Nehmen wir nochmal das Beispiel Chat-GPT4. Ihr habt was eingegeben bei Chat-GPT und bekommt folgenden Hinweis: „Es tut mir leid, aber ich kann Ihnen bei dieser Anfrage nicht helfen.“ Die GPT-4-Ausgabe verwendet Pronomen in der ersten Person, also Ich, was darauf hindeuten soll, dass man mit einer menschlichen Entität interagiert. Das ist eine bewusste Design-Entscheidung seitens OpenAI, die Ausgabe so zu formulieren. Diese Design-Entscheidung kann verschiedene Gründe haben.

Viele Automobilhersteller gestalten ihre Kühlergrills zum Beispiel so, dass sie menschlichen Gesichtern ähneln, um Käufer*innen anzusprechen. Einige Unternehmen bevorzugen ein „lächelndes“ Auto, um bei Verbraucher*innen wie freundliche Markenbotschafter zu wirken.

Einblendung Herbie

Quelle [19]

Das wird durch die Form des Kühlergrills erreicht. Andere Automarken entscheiden sich für einen „bösen Ausdruck“, um ihr Image von Aggression zu unterstreichen, also z. B. bei Poser Autos. Laut Landwehr, McGill und Herrmann entstehen Freundlichkeitswahrnehmungen aus dem Kühlergrill eines Autos, während Aggressivität sowohl durch den Grill als auch durch die Scheinwerfer vermittelt werden kann.

Quelle [20]

Anthropomorphe Sprache:

Anthropomorphe Sprache ist die Verwendung von menschenähnlichen Eigenschaften, Merkmalen oder Attributen zur Beschreibung nicht-menschlicher Wesen oder Phänomene. In dieser Sprache werden Tieren, Objekten oder abstrakten Konzepten oft Emotionen, Absichten oder Verhaltensweisen zugeschrieben, die normalerweise mit Menschen in Verbindung gebracht werden.

Vielleicht ist euch schon aufgefallen, dass auch in der KI-Forschung und Wissenschaft oft anthropomorphe Sprache verwendet wird.

Allein durch den Namen künstliche Intelligenz werden Erwartungen geweckt, indem Entitäten, die weder lebendig noch menschlich sind, eine menschliche Eigenschaft, nämlich Intelligenz, zugeschrieben wird, was implizite Annahmen über die Fähigkeiten von KI-Systemen offenbart. Die Verwendung einer solchen anthropomorphen Sprache ermutigt auch dazu, das Verhalten von Algorithmen als menschenähnlich zu interpretieren, was den Vergleich mit menschlichen Denkweisen erleichtert.

Quelle [21]

Take-Home Message

Anthropomorphismus, die Neigung, nicht-menschlichen Entitäten menschliche Eigenschaften zuzuschreiben, durchdringt unseren Alltag und hat auch Auswirkungen auf unsere Interaktionen mit Technologie. Wir haben auch die Tendenz, KI-Systeme zu vermenschlichen. Die bewusste Verwendung anthropomorpher Sprache und das Design von Technologie, um menschenähnlicher zu wirken, kann unser Verständnis von KI und unsere Erwartungen beeinflussen. Durch dieses Video soll euch vermittelt werden, dass es wichtig ist, sich dieser Tendenz bewusst zu sein und eine differenzierte Perspektive zu bewahren, um die Grenzen zwischen menschlicher Intelligenz und technologischer Funktionalität klar zu erkennen.

Quellen

Quelle [1][2][3][5][6]

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864-886.
<https://doi.org/10.1037/0033-295X.114.4.864>

Quelle [7] Culley, K. E., & Madhavan, P. (2013). A note of caution regarding anthropomorphism in HCI agents. *Computers in Human Behavior*, 29(3), 577-579.
<https://doi.org/10.1016/j.chb.2012.11.023>

Quelle [8] Li, M., & Suh, A. (2021). Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 4053-4062).
<https://doi.org/10.24251/HICSS.2021.493>

Quelle [9] Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494-1509. <https://doi.org/10.1016/j.chb.2007.05.007>

Quelle [10] Lee, J. G., Kim, K. J., Lee, S., & Shin, D. H. (2015). Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction*, 31(10), 682-691.
<https://doi.org/10.1080/10447318.2015.1070547>

Quelle [11] Niu, D., Terken, J., & Eggen, B. (2018). Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 28(6), 352-359. <https://doi.org/10.1002/hfm.20745>

Quelle [12] Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and

older adults. *Ergonomics*, 55(9), 1059-1072.
<https://doi.org/10.1080/00140139.2012.691554>

- Quelle [13] Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology*, 52, 113-117.
<https://doi.org/10.1016/j.jesp.2014.01.005>
- Quelle [14] Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology*, 85, Article 103870.
<https://doi.org/10.1016/j.jesp.2019.103870>
- Quelle [15] Mayer, M. M., Buchner, A., & Bell, R. (2023). Humans, machines, and double standards? The moral evaluation of the actions of autonomous vehicles, anthropomorphized autonomous vehicles, and human drivers in road-accident dilemmas. *Frontiers in Psychology*, 13, 1052729.
<https://doi.org/10.3389/fpsyg.2022.1052729>
- Quelle [16] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Quelle [17] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
- Quelle [18] Meineck, S. (2023, Juli 4). Olimpias Augen: Das Hype-Theater um moderne Chatbots. *netzpolitik.org*. <https://netzpolitik.org/2023/olimpias-augen-das-hype-theater-um-moderne-chatbots/>
- Quelle [20] Valenzuela, A., & Hadi, R. (2017). Implications of product anthropomorphism through design. In *The Routledge Companion to Consumer Behavior* (pp. 82-96). Routledge.
- Quelle [21] Placani, A. Anthropomorphism in AI: hype and fallacy. *AI Ethics* (2024).
<https://doi.org/10.1007/s43681-024-00419-4>

Bildmaterial

- Quelle [4] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Quelle [4] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *Proceedings of the 2021*

ACM Conference on Fairness, Accountability, and Transparency, 610–623.

<https://doi.org/10.1145/3442188.3445922>

- Quelle [4] Meineck, S. (2023, Juli 4). Olimpias Augen: Das Hype-Theater um moderne Chatbots. *netzpolitik.org*. <https://netzpolitik.org/2023/olimpias-augen-das-hype-theater-um-moderne-chatbots/>
- Quelle [4] Valenzuela, A., & Hadi, R. (2017). Implications of product anthropomorphism through design. In *The Routledge Companion to Consumer Behavior* (pp. 82-96). Routledge.
- Quelle [19] Placani, A. Anthropomorphism in AI: hype and fallacy. *AI Ethics* (2024). <https://doi.org/10.1007/s43681-024-00419-4>

Disclaimer

Transkript zu dem Video „KI und Ich : Anthropomorphisierung – Wenn KI vermenschlicht wird!“, Dr. Maike Meyer, Dr. Jacqueline Klusik-Eckert und Marie von Lobenstein M.A. Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.