Künstliche Intelligenz

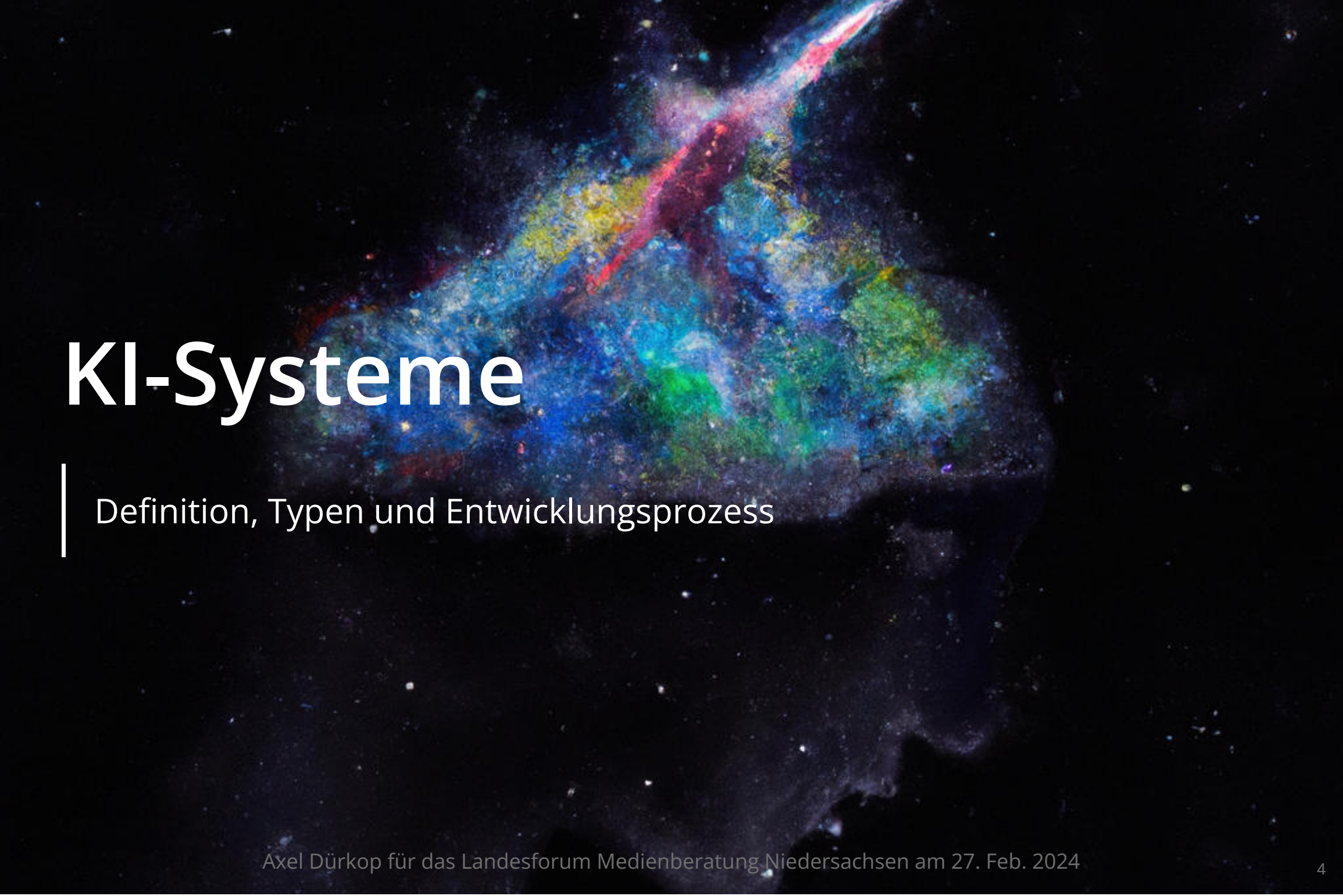
Chancen und Herausforderungen eines globalen sozio-technischen Systems

Axel Dürkop
me@axel-duerkop.de

Dienstag, 27. Februar 2024

Zentrale Fragen

- Was ist Künstliche Intelligenz (KI)?
- Warum sind KI-Systeme globale sozio-technische Systeme?
- Was sind die Chancen und Herausforderungen für Mensch und Gesellschaft?
- Was bedeutet KI für Bildung und Schule?



KI-Systeme

Definition, Typen und Entwicklungsprozess

Begriffsbestimmung KI

Enge KI

- Simulation menschlicher Fähigkeiten in einer Domäne
- Verwendung *maschinellen Lernens*
- spezifische Aufgaben oder Probleme

Breite KI

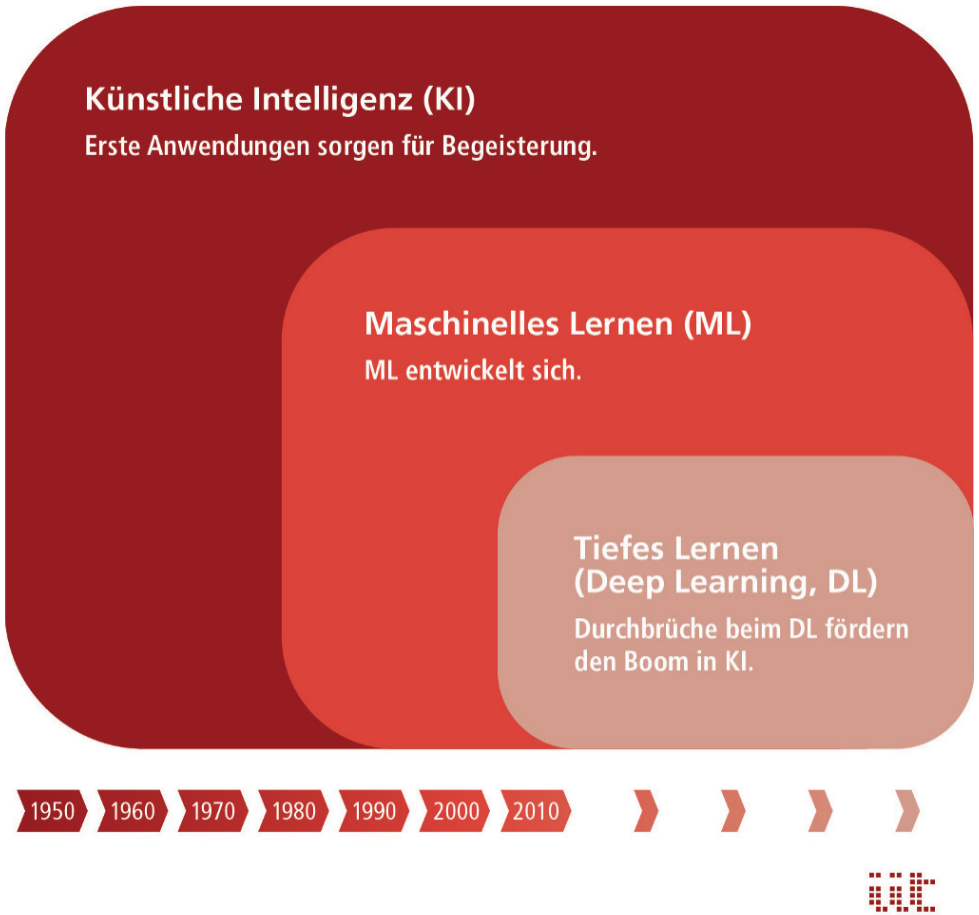
- Simulation menschlicher Fähigkeiten, nicht domänenspezifisch
- beschränkt auf sprachliche Ein- und Ausgabe

Starke KI

- (möglicherweise) perfekte Simulation menschlicher Kognition
- Vision: mentale Zustände, Einsichtsfähigkeit und Emotionen

Quelle: Deutscher Ethikrat (2023, S. 88)

Maschinelles Lernen



Methoden der KI (Kirste & Schürholz, 2019, S. 22)

Maschinelles Lernen umgibt uns in vielen Bereichen des täglichen (Arbeits)lebens.

- autonomes Fahren
- Robotik/industrielle Fertigung
- Überwachungssysteme, Militär
- Unternehmensanwendungen, E-Commerce
- Hochfrequenzhandel
- Medizin/Pflege
- Bildungsbereich

Generative KI-Systeme

- **Textgenerierende KI-Systeme**
  - bspw. ChatGTP, Gemini, Claude, Mistral und Llama 2 bekannt als Large Language Models (LLM)
- **Standbildgenerierende KI-Systeme**
  - bspw. Midjourney, DALLE oder Stable Diffusion
- **Bewegtbildgenerierende KI-Systeme**
  - bspw. Sora, Synthesia oder Pika
- **Tongenerierende KI-Systeme**
  - bspw. Stable Audio oder VALL-E

Entwicklungsprozesse von KI-Systemen

Alle Schritte finden auf Computerhardware statt.

Daten auswählen und sammeln

- Verwenden großer Datenmengen aus Internet oder privaten Quellen

Daten aufbereiten

- Annotieren der Daten, wenn möglich und sinnvoll, durch Menschen und Maschinen

Modelle trainieren (mit Grafikkarten (GPUs))

- Erlernen von Mustern, Klassen und Wahrscheinlichkeiten

Modelle evaluieren

- Überprüfen der Leistung durch Menschen und Maschinen

Sozio-technische Systeme

Sozio-technische Systeme implementieren Werte

Die Wirkung von KI-Systemen hängt von **Werten** ab, die bei ihrer Entwicklung und ihrem Einsatz leitend sind:

> "[...] AI systems are socio-technical systems. Their societal impact depends not only on the technology (data and algorithms) but also on the system's underlying goals and on the way an AI is embedded in an organisational structure." (Hallensleben et al., 2020, S. 10)

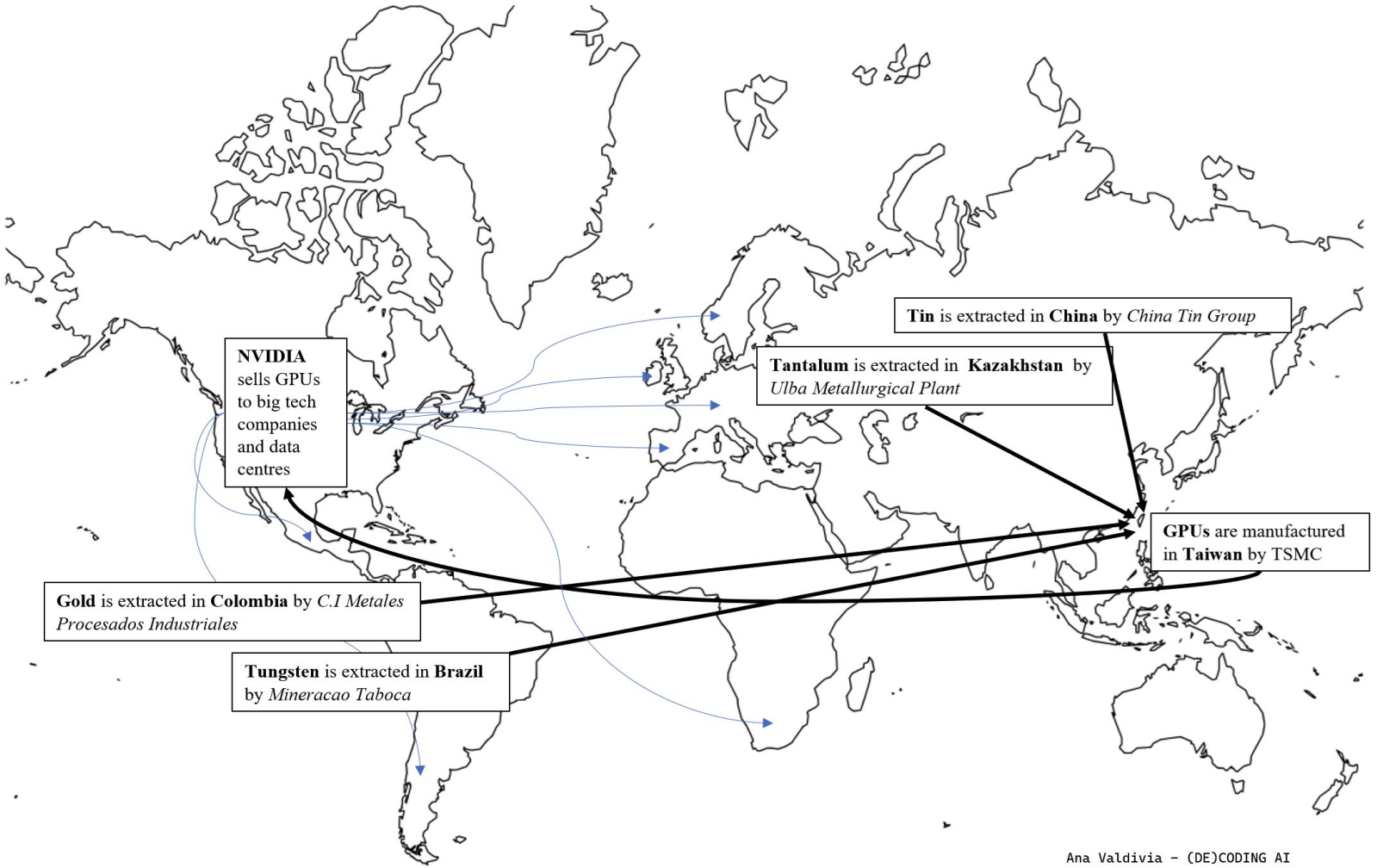
Die betrachteten Anwendungen wechselwirken mit der Organisation, der Kultur und der Gesellschaft, in der sie eingesetzt werden (Herrmann, 2003).

Auffassung der SWK (2023)

Auch die Ständige Wissenschaftliche Kommission der KMK weist implizit auf sozio-technische Dimensionen von KI hin:

> "Mit der Nutzung von LLM-Tools sind neben den Anforderungen an die Kompetenzen bei Lehrenden und Lernenden weitere Herausforderungen verbunden. Diese Herausforderungen sind vor allem **rechtlicher** und **ethischer Natur**, betreffen aber auch **Prüfungskultur** und **Qualitätssicherung**. Darüber hinaus sind weitere Fragen mit der Nutzung von LLM verbunden, die in diesem Impulspapier angesichts des Fokus auf Lehr-Lernprozesse nicht weiter behandelt werden sollen: **Politische und ethische Fragen**, die sich im Kontext der Programmierung stellen (z. B. der Umgang mit **Clickworkern**), aber auch Fragen des **Umweltschutzes** und **Energieverbrauchs**." (S. 15, Hervorh. A.D.)

Hardware, Energie, Wasser — und Menschen



Die Herstellung von Grafikkarten (GPUs) als Hardwaregrundlage von KI. Nicht abgebildet ist bspw. Agbogbloshie bei Accra, Ghana, wo ein Großteil des Elektroschrotts der westl. Welt (illegal) ausgeschlachtet wird. Quelle: Valdivia (2024).

Werteorientierter Reflexionsrahmen

Die 17 Nachhaltigkeitsziele der UN stellen einen möglichen, wenn auch nicht hinreichenden, Reflexionsrahmen dar, mit dem sich KI-Systeme in den Bereichen **Wirtschaft**, **Soziales** und **Umwelt** betrachten lassen.

Quelle: Bundesregierung (2018)

Werteorientierter Reflexionsrahmen

Die 17 Nachhaltigkeitsziele der UN stellen einen möglichen, wenn auch nicht hinreichenden, Reflexionsrahmen dar, mit dem sich KI-Systeme in den Bereichen **Wirtschaft**, **Soziales** und **Umwelt** betrachten lassen.

Auswahl im Folgenden behandelter SDGs. Quelle: Bundesregierung (2018)

Gesundheit und Wohlergehen

Ein gesundes Leben für alle Menschen jeden Alters gewährleisten und ihr Wohlergehen fördern.

Chancen

KI-Systeme leisten in der Medizin gute Dienste, wenn sie bspw. zur Unterstützung der Diagnose eingesetzt werden.

- **Zusammen besser:** Ärzt*innen und KI werden gemeinsam besser bei der Erkennung von Brustkrebs (Leibig et al., 2022)
- **Selbstvorsorge:** Smartphone-basierte KI-Systeme zur Hautkrebserkennung können die Früherkennung verbessern (Esteva et al., 2017)

Gesundheit und Wohlergehen

Ein gesundes Leben für alle Menschen jeden Alters gewährleisten und ihr Wohlergehen fördern.

Herausforderungen

KI-Systeme werden zunehmend für **Deepfakes** und **Pornografie** eingesetzt. Die Digitalisierung der Medizin basiert auf sehr persönlichen Daten, deren Schutz oft unzureichend ist.

- **Sexualisierte Gewalt:** Nicht-einvernehmliche Nutzung von Bildern mit Folgen für die mentale Gesundheit der Betroffenen (Heikkilä, 2024; Maiberg, 2023)
- **Datenschutz:** Bilder aus Krankenakte tauchen in KI-Trainingsdatensatz auf (Nordenbrock, 2022)

Bezahlbare und saubere Energie

Zugang zu bezahlbarer, verlässlicher, nachhaltiger und zeitgemäßer Energie für alle sichern.

Chancen

Maschinelles Lernen wird zunehmend für **datengetriebene Entscheidungsfindung** eingesetzt.

- **Wartung:** In der Kombination von Mensch, Drohne und KI wird eine Zukunft für die Wartung des britischen Stromnetzes gesehen (Ralston, 2022).
- **Steuerung:** Im Großen wie im Kleinen (Privathaushalte) kann KI bei der Steuerung des komplexen (erneuerbaren) Stromnetzes unterstützen (Kim, 2023).

Bezahlbare und saubere Energie

Zugang zu bezahlbarer, verlässlicher, nachhaltiger und zeitgemäßer Energie für alle sichern.

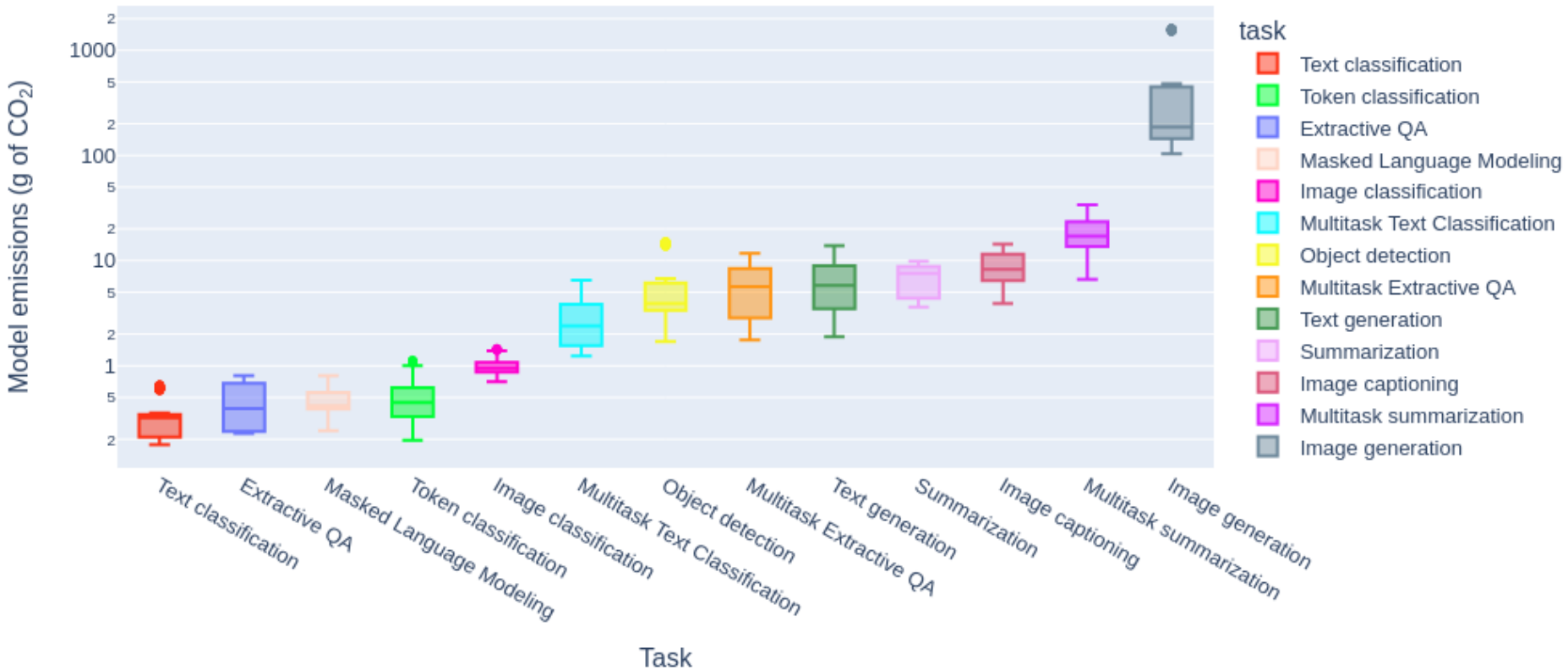
Herausforderungen

Derzeit verbrauchen Training und Betrieb von KI-Systemen **große Mengen Energie** und emittieren entsprechend viel **CO_2**.

- **KI-Standort und Energiepoltik:** Eine KI-Trainingsstunde in Indien (fossil) oder Kanada (erneuerbar) macht für den CO_2-Ausstoß einen Unterschied. (HuggingFace, 2021)
- **Generieren kostet:** Erste Studien weisen nach, welche Kosten für unterschiedliche Machine-Learning-Ansätze anfallen (Luccioni et al., 2023). Der Betrieb von Modellen addiert sich irgendwann auf zu den Kosten des Trainings.

Bezahlbare und saubere Energie

KI-Modelle verbrauchen je nach Funktionsumfang weniger oder mehr Strom und produzieren entsprechend auch CO_2.



Quelle: Luccioni et al. (2023, S. 1)

Nachhaltige Wassernutzung

Chancen

Maschinelles Lernen kann zum **Entdecken unbekannter Muster und Phänomene** eingesetzt werden, die Menschen sinnlich und kapazitiv nicht erfassen können.

- **Biodiversität:** KI zählt Fische und Plastikmüll und schützt die Meere (Elagali et al., 2022)
- **Ökologisches Gleichgewicht:** KI findet unbekannte Fischflotten und hilft, den Fischbestand zu kontrollieren (Paolo et al., 2024)

Nachhaltige Wassernutzung

Herausforderungen

- **Steigender Verbrauch:** Training und Betrieb von KI-Modellen verbrauchen große Mengen Wasser (Li et al., 2023)

> “So hat Microsoft publik gemacht, dass der Wasserverbrauch des Konzerns im vergangenen Jahr um ganze 34 Prozent (auf 6,4 Millionen m^3) in die Höhe gesprungen ist, bei Google gab es einen Anstieg um 20 Prozent.” (Holland, 2023)

Menschenwürdige Arbeit und Wirtschaftswachstum

Nachhaltiges Wirtschaftswachstum und menschenwürdige Arbeit für alle - dauerhaftes, breitenwirksames und nachhaltiges Wirtschaftswachstum, produktive Vollbeschäftigung und menschenwürdige Arbeit für alle fördern.

Herausforderungen

Extraktivismus bestimmt die Produktionskette der Digitalisierung.

- **Mineral Mining:** Für die Produktion von Hardware für die digitale Welt sind verschiedene Seltene Erden erforderlich, die weltweit geschürft werden (Crawford, 2022).
- **Data Mining:** Menschen in Ländern des globalen Südens arbeiten unter unwürdigen Bedingungen an der Datenqualität von KI-Systemen (Chen, 2019; Dzieza, 2023; Hao & Hernández, 2022; Marienhagen & Ohdah, 2023; Perrigo, 2023; Williams, 2022).

Hochwertige Bildung bezogen auf LLMs (SWK, 2023)

Inklusive, gleichberechtigte und hochwertige Bildung gewährleisten und Möglichkeiten des lebenslangen Lernens für alle fördern.

Chancen

- Entwicklung und Anpassung von Unterrichtsmaterial (vgl. S. 8)
- Unterstützung bei Kreativitätsprozessen und Recherche (vgl. S. 10)
- Unterstützung bei der Textproduktion (vgl. S. 11)
- Förderung von Individualisierung und Binnendifferenzierung (vgl. S. 11)
- summative und formative Beurteilung (vgl. S. 12)
- **Entlastung von Lehrenden durch neue Aufgabenaufteilung denkbar?**

Hochwertige Bildung bezogen auf LLMs (SWK, 2023)

Inklusive, gleichberechtigte und hochwertige Bildung gewährleisten und Möglichkeiten des lebenslangen Lernens für alle fördern.

Herausforderungen

- Nutzer*innen müssen die “[...] Kompetenz zur Bewertung, Korrektheit und Vertrauenswürdigkeit der Inhalte besitzen [...]” (S. 9)
- lernförderliche Aktivitäten könnten an KI-Systeme delegiert werden (vgl. S. 10f.)
- “Motivationsverluste sind zu befürchten, wenn Lernende den Eindruck haben, Aktivitäten ausüben zu müssen, die eine KI übernehmen könnte.” (S. 10)
- LLMs ersetzen nicht Lernbegleitung durch Lehrkräfte, da sie pädagogische Ziele nicht kennen (vgl. S. 12)
- neue Prüfungsformate sind notwendig (S. 16ff.)

Versierte Koaktivität

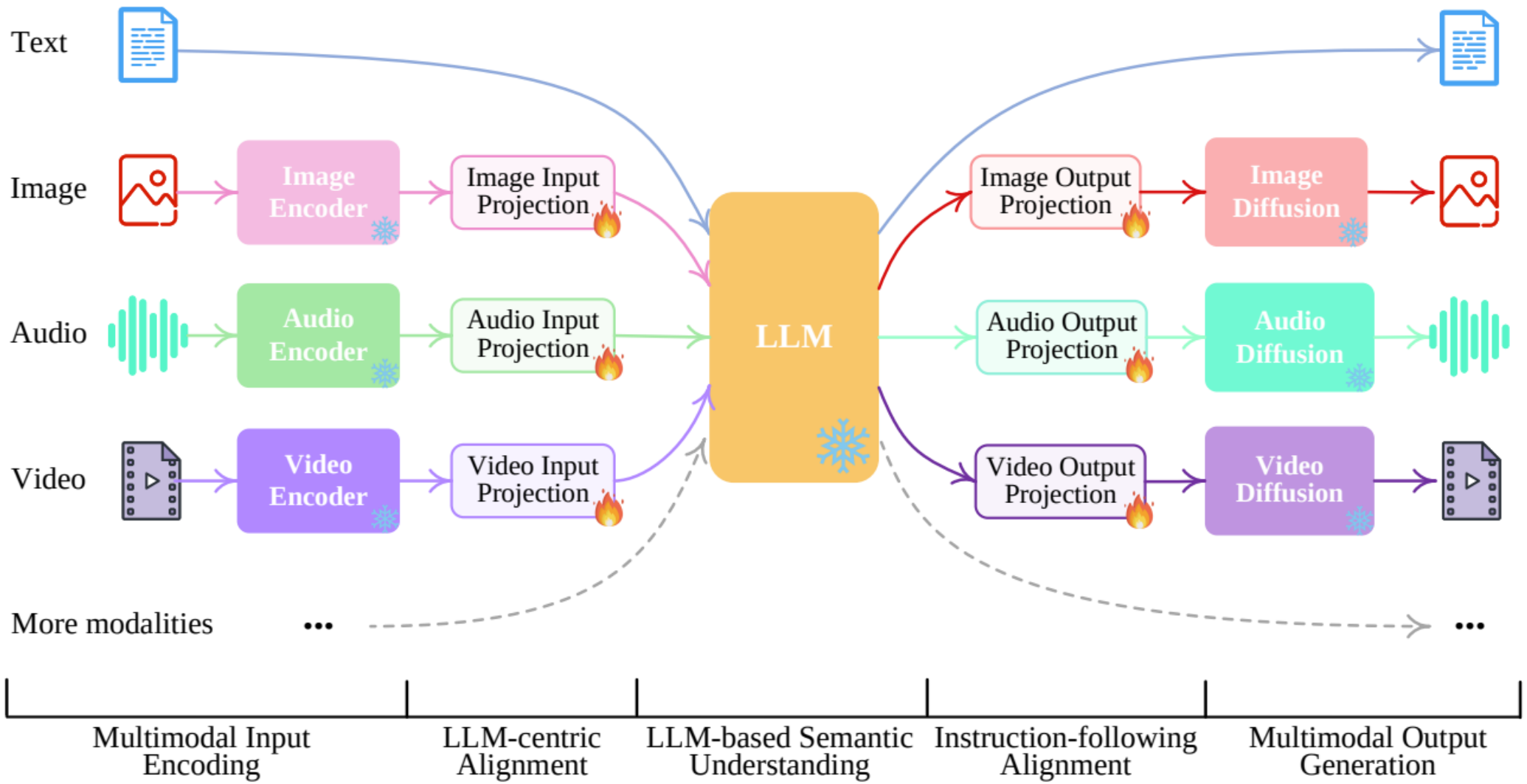
Versierte Koaktivität gestalten

Die SWK spricht von **“versierter Koaktivität”** (S. 4) als Nutzungsziel von LLMs durch Lernende und nennt diese eine **“wichtige Zukunftskompetenz”** (S. 16).

> “Erst wenn Lernende bereichsspezifisches Vorwissen mitbringen, verstanden haben, wie LLM trainiert wurden, wie die automatisierte Textproduktion funktioniert, zu welchem Zweck LLM aufgesetzt wurden und welche Rolle Prompts für das Antwortverhalten von LLM spielen, können sie eine kompetente Einordnung der durch LLM gelieferten Informationen vornehmen und diese für das eigene Lernen produktiv nutzen.” (S. 13)

Was bedeutet “versierte Koaktivität” für Lehrkräfte?
Sind KIs die neuen Kolleg*innen?

Nicht nur auf LLM beschränkt



Design eines multimodalen LLMs (MM-LLM) von Wu et al. (2023, S. 1), das unterschiedliche Medienarten aufnehmen und ausgeben kann. Zur Energiebilanz von MM-LLMs vgl. Luccioni et al. (2023)

Fazit und Ausblick

Chancen und Herausforderungen

- **KI-Systeme**
  - können als sozio-technische Systeme begriffen werden
  - implementieren absichtlich oder unabsichtlich Werte
  - werfen zahlreiche lernhaltige Fragen auf, die in Schule und Beruf mitgestaltet werden können
  - sollten bei Entwicklung und Nutzung zusätzlich aus einer nicht-technischen Perspektive begleitet werden
  - sind eine gestalterische und regulatorische Herausforderung für Politik, Gesellschaft und Individuum

"Technical design in accordance with human values is increasingly considered a building block for shaping the digital future." (Jacobs et al., 2021)

Kontakt

https://axel-duerkop.de
me@axel-duerkop.de
@xldrkp@scholar.social

Mediennachweise

- “Chuquicamata-Kupfermine in Chile”: Foto von Alexander Gerst auf flickr, CC BY-SA 2.0 DEED
- “Dart”: Foto von Ricardo Arce auf Unsplash
- “Nebula Head” auf “Was ist KI?”: DALL-E 2, ” artificial intelligence in the implosion of a nebula, watercolor, dark background”
- “Entwicklungsprozesse von KI-Systemen”: Photo by Eaters Collective on Unsplash
- “Sozio-technische Systeme”: DALL-E in ChatGPT 4.0 (OpenAI) mit dem folgenden Prompt “A digital globe or network with a more rugged and textured look, connecting dots across continents in a landscape orientation. The image should depict the global scale of socio-technical systems, emphasizing a less polished, more raw and organic feel in a 16:9 aspect ratio. The globe should be at the center, surrounded by network lines or data streams connecting different points in a slightly chaotic manner. Silhouettes of people interacting through devices like laptops, smartphones, and tablets should look a bit rough around the edges, suggesting a more grassroots level of digital communication. The background should be darker with a gritty texture, making the digital connections and the globe stand out in a landscape format. This version should convey a sense of global digital collaboration and the integration of social and technical elements in a more unrefined, authentic manner.”
- “Handshake” auf “Versierte Koaktivität”, Photo by Andy Kelly on Unsplash
- “Fazit und Ausblick”: Photo by Patrick Perkins on Unsplash
- “Sternenhimmel in Orange” auf “Kontakt”, Foto von Sid Suratia auf Unsplash

Nachnutzung

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz . Diese Lizenz erlaubt unter Voraussetzung der Namensnennung des Urhebers die Bearbeitung, Vervielfältigung und Verbreitung des Materials in jedem Format oder Medium für beliebige Zwecke, auch kommerziell, sofern das neue entstandene Werk unter derselben Lizenz wie das Original verbreitet wird.

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z. B. Schaubilder, Abbildungen, Fotos, Textauszüge und Medien, die in den Mediennachweisen gelistet sind, erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Acknowledgments

Die Keynote, für die die Folien erstellt wurden, fand am 27. Februar 2024 im Rahmen des Landesforums der Medienberatung Niedersachsen in Soltau statt. Die Entwicklung des Vortrags wurde finanziert durch das Niedersächsische Landesinstitut für schulische Qualitätsentwicklung (NLQ).

Referenzen

Bundesregierung. (2018). *Agenda 2030: Unsere Nachhaltigkeitsziele* [Homepage der Bundesregierung]. Die Bundesregierung informiert | Startseite. https://www.bundesregierung.de/breg-de/themen/nachhaltigkeitspolitik/nachhaltigkeitsziele-erklaert-232174

Chen, A. (2019, Mai 13). *How Silicon Valley's Successes Are Fueled by an Underclass of „Ghost Workers"*. The Verge. https://www.theverge.com/2019/5/13/18563284/mary-gray-ghost-work-microwork-labor-silicon-valley-automation-employment-interview

Crawford, K. (2022). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Deutscher Ethikrat. (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz* [Stellungnahme]. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

Dzieza, J. (2023, Juni 20). *Inside the AI Factory*. Intelligencer. https://nymag.com/intelligencer/article/ai-artificial-intelligence-humans-technology-business-factory.html

Elagali, A., Robidart, J., Quigley, K. M., & Bayer, P. (2022, Dezember 14). *Spotting Plastic Waste from Space and Counting the Fish in the Seas: Here's How AI Can Help Protect the Oceans*. The Conversation. http://theconversation.com/spotting-plastic-waste-from-space-and-counting-the-fish-in-the-seas-heres-how-ai-can-help-protect-the-oceans-196222

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*, *542*(7639, 7639), 115–118. https://doi.org/10.1038/nature21056

Hallensleben, S., Hustedt, C., Fetic, L., Fleischer, T., Grünke, P., Hagendorff, T., Hauer, M., Hauschke, A., Heesen, J., Herrmann, M., Hillerbrand, R., Hubig, C., Kaminski, A., Krafft, T., Loh, W., Otto, P., & Puntschuh, M. (2020). *From Principles to Practice. An Interdisciplinary Framework to Operationalise AI Ethics*. Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/from-principles-to-practice-wie-wir-ki-ethik-messbar-machen-koennen

Hao, K., & Hernández, A. P. (2022). 90 Cent pro Stunde (G. Honsel, Übers.). *MIT Technology Review*, *06*, 34–40.

Heikkilä, M. (2024, Januar 31). *Der Fall Taylor Swift: Drei Möglichkeiten, gegen Deepfake-Pornografie vorzugehen* [Nachrichtenseite]. heise online. https://www.heise.de/hintergrund/Der-Fall-Taylor-Swift-Drei-Moeglichkeiten-gegen-Deepfake-Pornografie-vorzugehen-9613730.html

Herrmann, T. (2003). Learning and Teaching in Socio-technical Environments. *Informatics and the Digital Society: Social, Ethical and Cognitive Issues*, 59–71. https://doi.org/10.1007/978-0-387-35663-1_6

Holland, M. (2023, September 11). *Wohl wegen ChatGPT: Wasserverbrauch von Microsoft 2022 um ein Drittel gestiegen*. heise online. https://www.heise.de/news/Wohl-wegen-ChatGPT-Wasserverbrauch-von-Microsoft-2022-um-ein-Drittel-gestiegen-9300486.html

HuggingFace (Regisseur). (2021, November 15). *The Carbon Footprint of Transformers*. https://www.youtube.com/watch?v=ftWlj4FBHTg

Jacobs, M., Kurtz, C., Simon, J., & Böhmann, T. (2021). Value Sensitive Design and Power in Socio-Technical Ecosystems. *Internet Policy Review*, *10*(3). https://doi.org/10.14763/2021.3.1580

Kim, J. (2023, November 22). *Four Ways AI Is Making the Power Grid Faster and More Resilient* [Nachrichten aus der Wissenschaft]. MIT Technology Review. https://www.technologyreview.com/2023/11/22/1083792/ai-power-grid-improvement/

Kirste, M., & Schürholz, M. (2019). Einleitung: Entwicklungswege zur KI. In V. Wittpahl (Hrsg.), *Künstliche Intelligenz* (S. 21–35). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58042-4_1

Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K., & Umutlu, L. (2022). Combining the Strengths of Radiologists and AI for Breast Cancer Screening: A Retrospective Analysis. *The Lancet Digital Health*, *4*(7), e507–e519. https://doi.org/10.1016/S2589-7500(22)00070-X

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023, April 6). *Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models*. https://doi.org/10.48550/arXiv.2304.03271

Luccioni, A. S., Jernite, Y., & Strubell, E. (2023, November 28). *Power Hungry Processing: Watts Driving the Cost of AI Deployment?* https://doi.org/10.48550/arXiv.2311.16863

Maiberg, E. (2023, August 22). *Inside the AI Porn Marketplace Where Everything and Everyone Is for Sale*. 404 Media. https://www.404media.co/inside-the-ai-porn-marketplace-where-everything-and-everyone-is-for-sale/

Marienhagen, N., & Ohdah, D. (2023, Juli 16). Training für künstliche Intelligenz: So sieht die Arbeit von Clickworkern aus. *Der Spiegel: Netzwelt*. https://www.spiegel.de/netzwelt/web/training-fuer-kuenstliche-intelligenz-so-sieht-die-arbeit-von-clickworkern-aus-a-7c018325-440d-461c-a004-6b3b35fb9ed1

Nordenbrock, K. (2022, September 26). *KI-Künstlerin findet Bilder aus eigener Krankenakte in KI-Trainingssat wieder*. t3n Magazin. https://t3n.de/news/bilder-aus-privater-krankenakte-in-ki-trainingssatz-1500514/

Paolo, F. S., Kroodsma, D., Raynor, J., Hochberg, T., Davis, P., Cleary, J., Marsaglia, L., Orofino, S., Thomas, C., & Halpin, P. (2024). Satellite Mapping Reveals Extensive Industrial Activity at Sea. *Nature*, *625*(7993 7993), 85–91. https://doi.org/10.1038/s41586-023-06825-8

Perrigo, B. (2023, Januar 18). *Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer*. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/

Ralston, W. (2022, Juli 5). Autonomous Drones Could Soon Run the UK's Energy Grid. *Wired*. https://www.wired.com/story/autonomous-drones-could-soon-run-the-uks-energy-grid/

SWK (Hrsg.). (2023). *Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz*. 31. https://doi.org/10.25656/01:28303

Valdivia, A. (2024). *The Supply Chain Capitalism of AI: A Call to (Re)Think Algorithmic Accountability, Harms and Resistance.*

Williams, A. (2022). *The Exploited Labor Behind Artificial Intelligence*. https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence