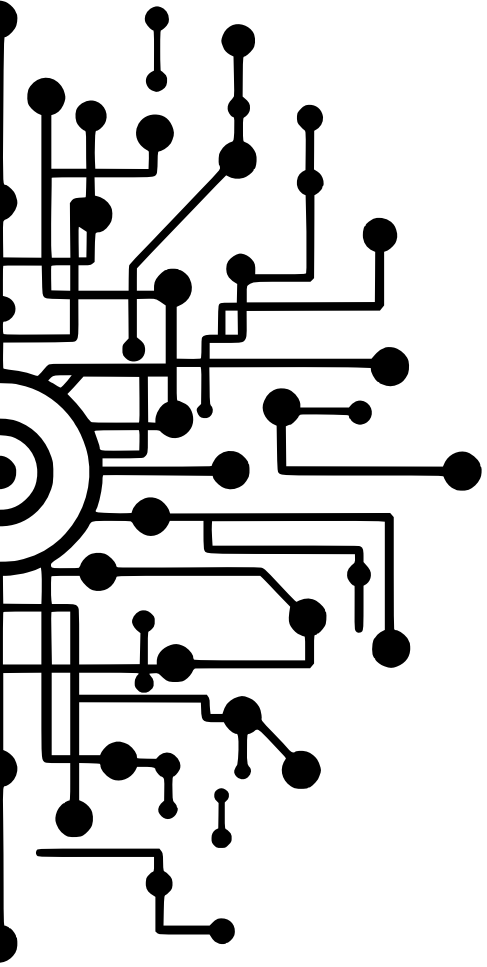


WoLKE

(wissenschaftliche) Gütekriterien & Ethik

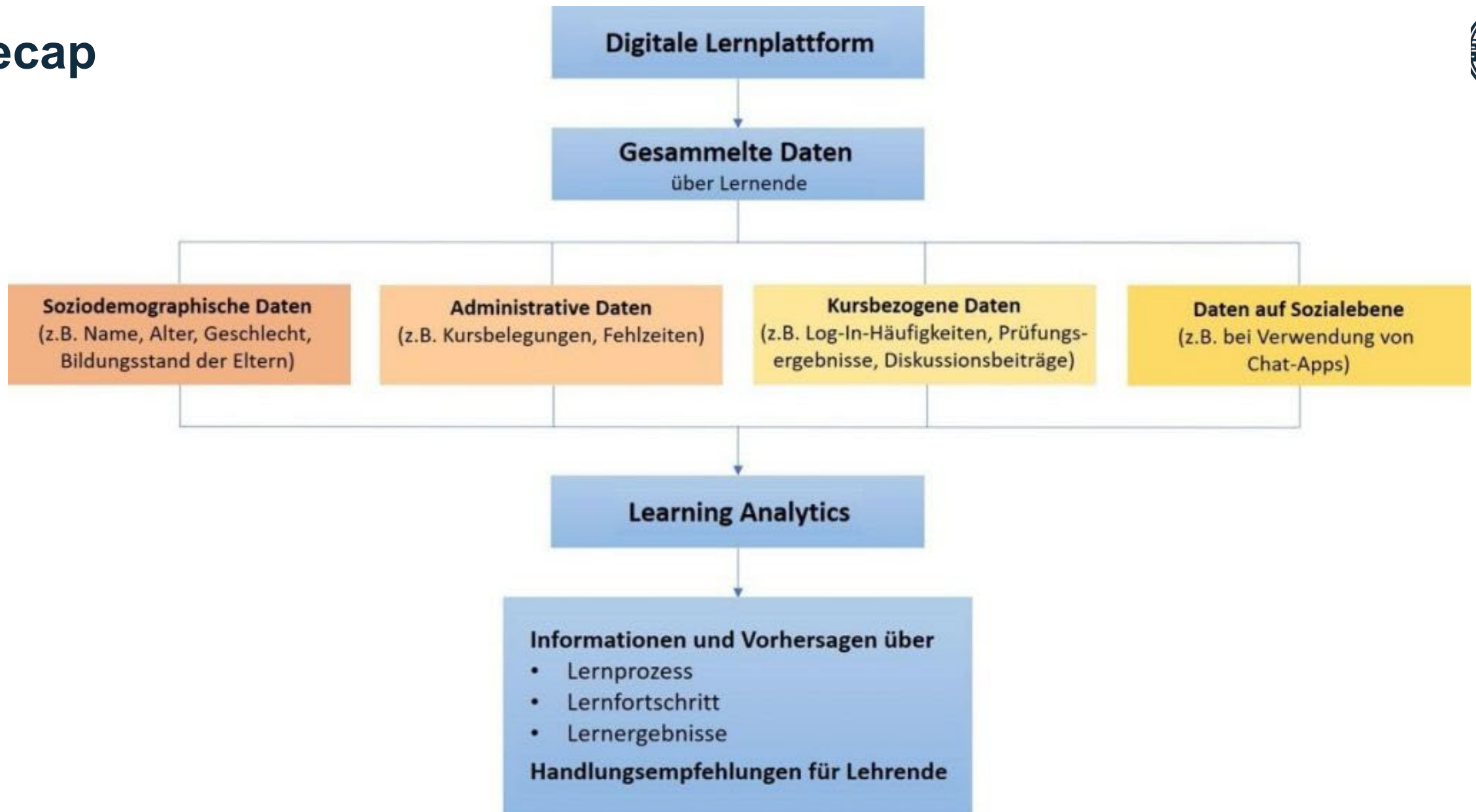


Semesterübersicht



Nr.	Module	Inhalte
1	Einstieg	Organisatorisches
2	Technologische Grundlagen	Algorithmen
3		Maschinelles Lernen
4		Neuronale Netze & Deep Learning
5		LLMs & GPTs
6	Umgang mit KI	Learning Analytics
7		Gütekriterien & Ethik
8		Tool-Erkundung
9		Prompt-Engineering
10	Praxis: KI & Schule	Didaktisches Grundwissen
11		Inhalte erstellen
12		Peer Feedback
13		Prüfungen im Kontext von KI
14	Abschluss	Abschlussevaluation & -reflexion

Recap



Learning Analytics: Die digitale Zukunft des Lernens – Netzwerk Digitale Bildung via www.netzwerk-digitale-bildung.de

Sitzungsziele – Gütekriterien & Ethik



Am Ende dieser Sitzung können die Studierenden...

- Beispiele für *ethische Konflikte* der Verwendung von KI (im Unterricht) **nennen**
- Den *Interpretability-Accuracy Trade-off* **beschreiben**
- Die Begriffe *Kausalität* und *Korrelation* anhand eines Beispiels **erklären**
- Die *Funktionsweise mathematischer Logik mit probabilistischer KI* **vergleichen** und unter verschiedenen Gesichtspunkten (z. B. wissenschaftlich, bildungspolitisch, rechtlich, ...) **diskutieren**



Warum machen wir das?

Ethische Überlegungen sind notwendig, um KI nicht nur anzuwenden, sondern ihre Ergebnisse, Grenzen und Folgen kritisch beurteilen zu können.

Einstieg

Bitte lest euch diesen Artikel durch.

Ihr findet den Link zum Artikel, sowie das zitierte Paper in moodle

[ki-überzeugt-uns-besser-als-menschen-und-das-sollte-angst-machen](#)



The screenshot shows a LinkedIn post from Till Oberhummer, dated May 28, 2025. The post title is "KI überzeugt uns besser als Menschen – und das sollte uns Angst machen". The text of the post discusses a study by Mira Murati (former CTO of OpenAI) and a study from EPFL Lausanne, both indicating that AI is more persuasive than humans, especially when it has personal information about the user. The post includes a link to the article and a note that the paper is available in Moodle.

LinkedIn

Top-Inhalte Personen Learning Jobs Spiele

KI überzeugt uns besser als Menschen – und das sollte uns Angst machen

Till Oberhummer
Veröffentlicht: 28. Mai 2025

+ Folgen

Eine neue Studie zeigt: Wenn Maschinen wissen, wer wir sind, manipulieren sie uns 64% häufiger als andere Menschen. Willkommen in der Intention Economy.

Was Mira Murati, ehemalige CTO von OpenAI, bereits 2024 als "scary" bezeichnete, ist jetzt wissenschaftlich belegt: **Künstliche Intelligenz ist uns Menschen in der Überzeugungsarbeit überlegen – und zwar dramatisch.**

Eine bahnbrechende Studie der renommierten EPFL Lausanne, veröffentlicht in *Nature Human Behaviour*, bringt erschreckende Zahlen ans Licht: **GPT-4 überzeugt Menschen in 64,4% der Fälle erfolgreicher als menschliche Debattierer** – allerdings nur, wenn die KI persönliche Informationen über ihr Gegenüber besitzt.

[ki-überzeugt-uns-besser-als-menschen-und-das-sollte-angst-machen](#) via [www.linkedin.com](#)



Andere Länder kritischer

Nur jeder vierte Deutsche überprüft KI-Ergebnisse

Stand: 02.05.2025 14:01 Uhr

KI-Chatbots werben mit smarten Antworten auf Knopfdruck. Doch nicht immer sind die Angaben richtig. Dennoch überprüft hierzulande offenbar nur etwa jeder Vierte die Ergebnisse.

Nur gut ein Viertel der Nutzer in Deutschland überprüft einer Umfrage zufolge die Ergebnisse, die KI-Chatbots wie ChatGPT, Google Gemini und Microsoft Copilot erstellen. Konkret gaben 27 Prozent der Befragten hierzulande an, KI-generierte Texte, Bilder oder Übersetzungen gegenzuprüfen. Das zeigt die [internationale Analyse](#) der Prüfungs- und Beratungsgesellschaft EY zum Thema Künstliche Intelligenz (KI). Der Durchschnitt in den 15 untersuchten Länder lag bei 31 Prozent.

Screenshot [ki-blindes-vertrauen](#) via [www.tagesschau.de](#)

MINT vs KI

Die Logik als Fundament der Mathematik



Um *mathematisch korrekt* zu sein, müssen alle mathematischen Aussagen, Beziehungen, Beweise, etc. auch *logisch korrekt* sein.

→ Das heißt, dass logische Schlussregeln befolgt werden müssen und Kalküle der formalen Logik nicht verletzt werden dürfen.

Die klassische Logik (Aussagenlogik, Prädikatenlogik) ist deterministisch:

- Jede Aussage ist entweder wahr oder falsch.
- Der Wahrheitswert jeder zusammengesetzten Aussage lässt sich eindeutig aus den Wahrheitswerten ihrer Teilaussagen ableiten.
- Mithilfe von *Schlussregeln* kann von mehreren wahren Prämissen auf die Wahrheit einer bestimmten Konklusion geschlossen werden.

→ Diese Strukturen sichern die Nachvollziehbarkeit, Eindeutigkeit und Exaktheit mathematischer Argumentation.

Mathematik und die Funktionsweise generativer KI

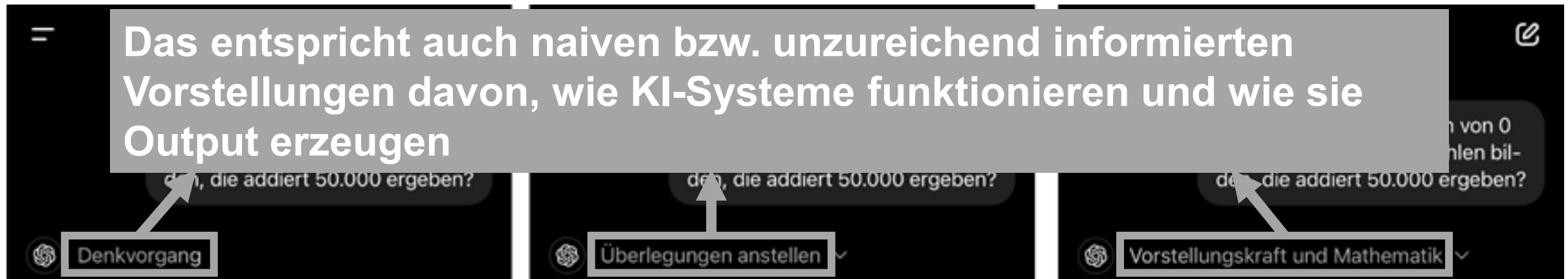


Mathematisches Arbeiten erfordert *logisches Denken*.

Generative KI-Tools, insbesondere Large Language Models (LLMs), werden immer verbreiteter in der mathematischen Bildung eingesetzt.

Aber "denken" diese KI-Systeme *logisch*?

Sie versuchen zumindest den Anschein zu erwecken:



Generative KI – probabilistische Funktionsweise



- LLMs sind generative KI-Modelle und arbeiten nicht mit formaler Logik, sondern erkennen Muster und berechnen Häufigkeiten des Auftretens von Wörtern (bzw. Tokens) in Konstellation mit anderen Wörtern (bzw. Kontext).
- Basierend darauf synthetisieren sie mittels stochastischer Algorithmen neue Ausgaben, um sprachlich plausible, aber nicht immer logisch korrekte Antworten zu generieren. (McCoy et al., 2023)
- Pseudozufällige Abweichungen sorgen für Variabilität und kreativ bzw. „menschlich“ wirkende Antworten. (McCoy et al., 2023; Vadhan, 2012)

Was LLMs *nicht* leisten können

LLMs haben **kein Verständnis von Inhalten oder Bedeutungen**,

- sondern nur davon, wie häufig ein bestimmtes Token in Konstellation mit bestimmten anderen Tokens in der dem Modell zugrundeliegenden Datenbasis (bestehend aus riesigen Mengen von Texten) vorkommt.

Sie haben z. B. kein Verständnis davon, dass ...

- 7 eine ungerade Zahl ist.
- die Menge der Quadrate eine Teilmenge der Rechtecke ist.
- 12 ein Vielfaches von 1, 2, 3, 4, 6 oder 12 ist.

LLMs **wenden keine logischen Schlussregeln an**, um die formale Korrektheit und Konsistenz ihrer Ausgaben sicherzustellen oder zu überprüfen,

- sondern bauen ihre Ausgaben basierend darauf auf, welche Tokenfolgen besonders wahrscheinlich sind (ohne jedoch deterministisch immer das wahrscheinlichste nachfolgende Token zu wählen).
- **Dennoch erzeugen sie mathematisch überwiegend korrekte Ausgaben!**

GPT 4 besteht das bayerische Mathe-Abitur



Januar 2023

Mai 2023

Fach	Abi 2022 (GPT 3.5)	Abi 2023 (GPT 4)
Deutsch	3	10
Mathematik	4	10
Informatik	2	11
Ethik	4	10
Geschichte	9	11

[chatgpt-ki-besteht-bayerisches-abitur-mit-bravour](https://www.br.de) via www.br.de



Wo liegt denn dann das Problem?

Kausalität

vs

Korrelation



Kausalität bedeutet:

Eine Ursache bewirkt eine bestimmte Wirkung.

Beispiel aus der Physik:

Wenn ich die Kraft auf einen Körper verdopple,
verdoppelt sich auch seine Beschleunigung.
Ursache: Kraft; Wirkung: Beschleunigung

→ **kausaler Zusammenhang** ($F = m \cdot a$)

Zentrale Eigenschaften kausaler Modelle:

- Sie beruhen auf **Naturgesetzen, mathematischen Beweisen** oder **logisch hergeleiteten Zusammenhängen**.
- Sie ermöglichen **gezielte Vorhersagen und technische Steuerung**.
- Sie sind in Experimenten **wiederholbar und überprüfbar**.

Korrelation bedeutet:

Zwei Dinge treten gemeinsam auf – aber nicht zwangsläufig, weil eins das andere verursacht.

Beispiel aus der Datenanalyse:

Je mehr Eis verkauft wird, desto mehr Menschen ertrinken im Sommer.

→ **Kein kausaler Zusammenhang!**

(Beide Phänomene steigen, weil es **heißer ist**.)

Für KI bedeutet das:

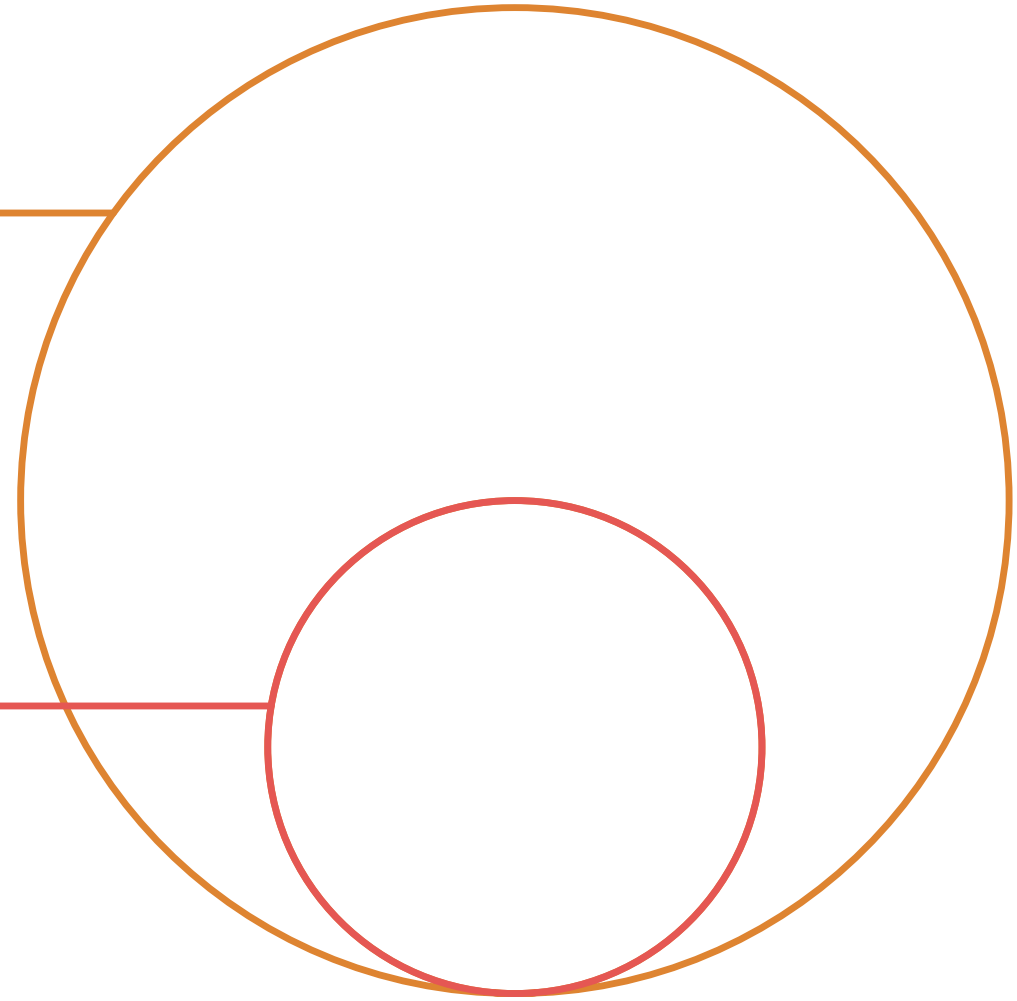
- Sie erkennt: *Wenn X vorkommt, dann häufig auch Y.*
- Aber sie weiß nicht: *Warum? Oder: Was ist Ursache, was ist Folge?*
- Deshalb entstehen oft **Fehlinterpretationen**, wenn man KI-Ergebnisse **als kausale Wahrheiten missversteht**.

Korrelation

eine Beziehung zwischen zwei oder mehreren Merkmalen, Zuständen oder Funktionen.

Kausalität

Ursache-Wirkungs-Beziehung zwischen Variablen



**Wenn KI auf
Wissenschaft trifft...**

Trifft KI auf Wissenschaft

...











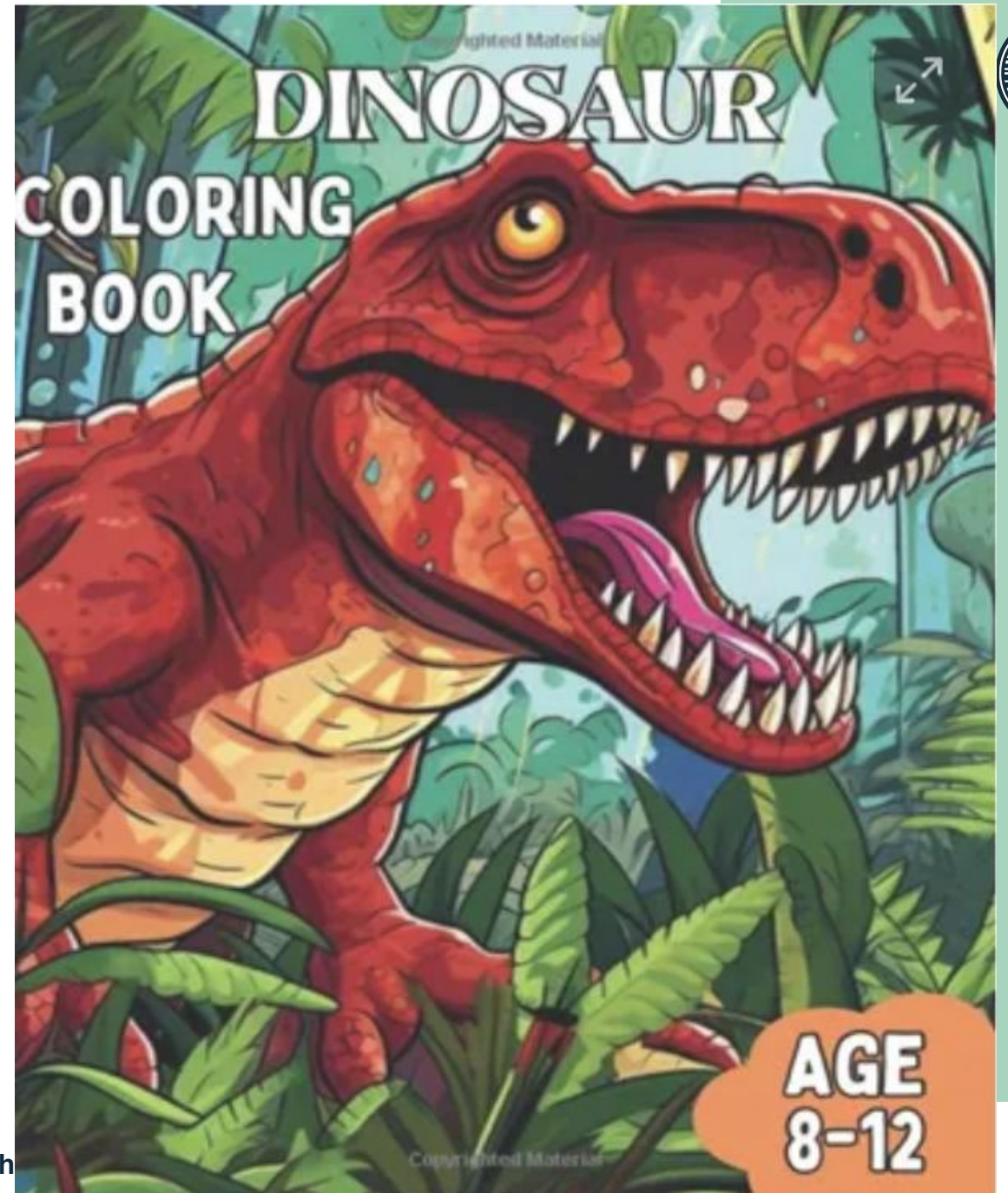
**Namensstreit
in der Taxonomie**

Ethische Aspekte

Ethischer Aspekt: Die Integrität der Wissenschaft

„ *'passives Einkommen' - durch
KI generierte (Kinder-) Bücher*

AI Is Coming For Your Children -
by The Only Robert Evans



Ethischer Aspekt: Die Integrität der Wissenschaft



“*Google AI just might kill you: it misidentified a Destroying Angel mushroom as an edible "button mushroom"*”

↖
Zucht-Champignon



[Google AI just might kill you: it misidentified a Destroying Angel mushroom as an edible "button mushroom" : r/technology](#)

Ethischer Aspekt: Die Integrität der Wissenschaft



r/LegalAdviceUK • vor 3 Monaten
Virtual_Cellist_736



Family poisoned after using AI-generated mushroom identification book we bought from major online retailer.



*Unfortunately, the book in question was not accurate. A closer investigation reveals that the **images of mushrooms are AI generated**, and we have now found two instances of text where a sentence ends and is followed up with a random questions or fourth-wall breaking statements.*

For example:

*"In conclusion, morels are delicious mushrooms which can be consumed from August to the end of Summer. **Let me know if there is anything else I can help you with.**"*



[Family poisoned after using AI-generated mushroom identification book we bought from major online retailer. : r/LegalAdviceUK](#)

Ethischer Aspekt: Betrug & DeepFakes

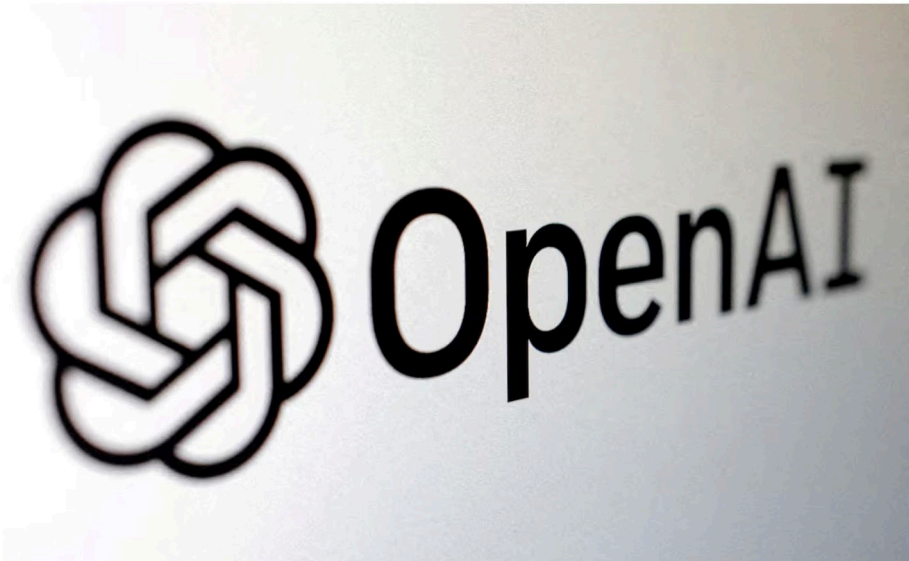


***Absprachen und Deals
zwischen KI-Entwicklern und
Dritten beeinflussen den
generierten Output***

Global news publisher Axel Springer partners with OpenAI in landmark deal

By Helen Coster

December 13, 2023 7:49 PM GMT+1 · Updated December 13, 2023



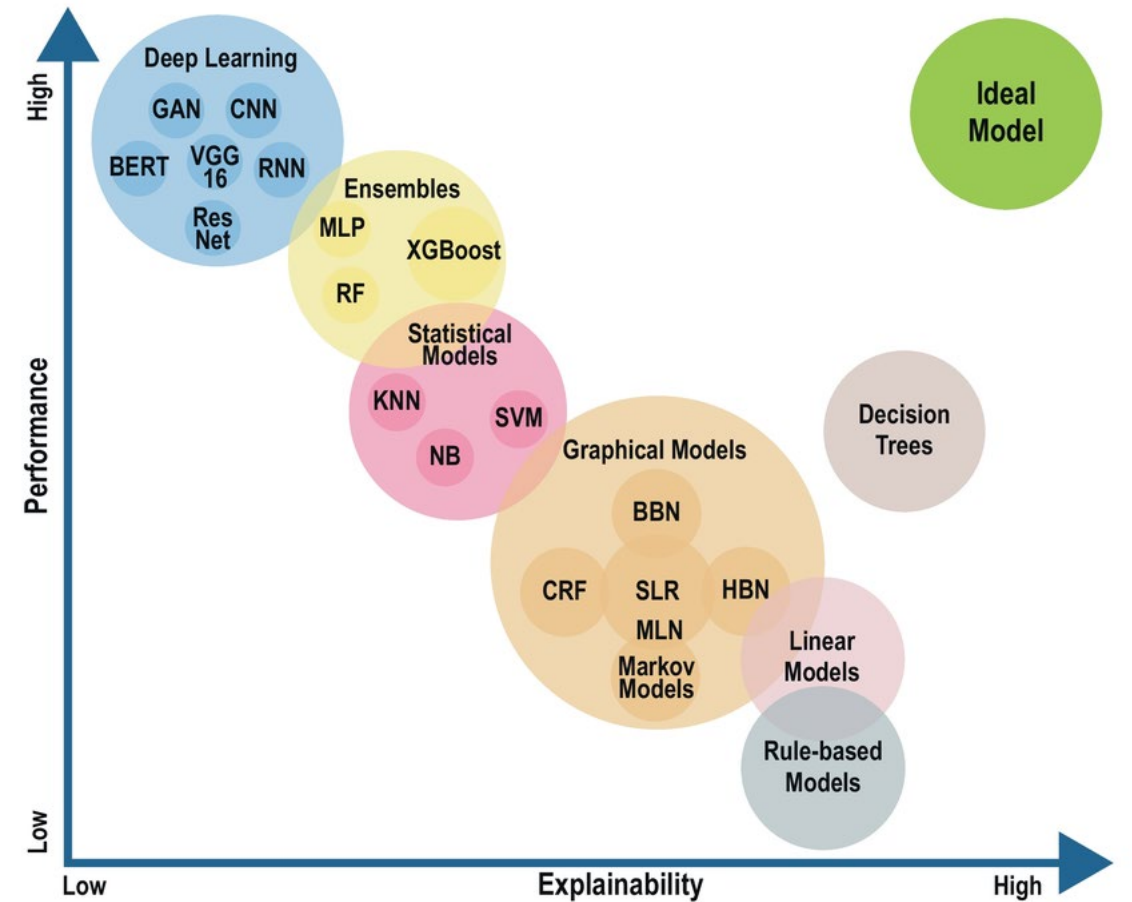
[global-news-publisher-axel-springer-partners-with-openai-landmark-deal](https://www.reuters.com/global-news-publisher-axel-springer-partners-with-openai-landmark-deal) via www.reuters.com

Der Interpretability-Accuracy Trade-off



Je präziser und komplexer ein Modell wird, desto geringer ist in der Regel seine Erklärbarkeit;

Je einfacher und besser erklärbar ein Modell ist, desto häufiger muss man Abstriche bei der Genauigkeit des Outputs machen.



[A Systematic Review. Cognitive Computation](#) via [Viswan](#)



Das Black-Box Spiel



Gruppenarbeit



15 min

Wir spielen heute **“Black Box”**.

Dort begegnet ihr einer KI, die euch eine Aufgabe stellt. Eure Aufgabe ist es herauszufinden, wie diese KI funktioniert (*findet das Muster*)



1. **Bildet Gruppen** aus *mindestens 4* Personen
2. **Wählt** in jeder Gruppe **eine** (mutige) freiwillige Person, die unsere **KI** spielt > **Alle anderen** Teammitglieder **sind die “Nutzer”**
3. **Aufgabe der Nutzer:**
 - Findet durch geschicktes Fragen/ Antworten heraus, wie eure KI funktioniert
 - Wenn du meinst die richtige Lösung herausgefunden zu haben, dann **darfst du niemandem die Lösung verraten, aber**
 - Deinen Teammitgliedern durch gezieltes Antworten / Interagieren mit der KI, Hinweise auf die richtige Lösung geben.
4. **Nach Ablauf der Zeit** können alle Nutzer Vermutungen abgeben, wie die KI programmiert wurde
5. **Anschließend** darf die KI ihre Programmierung offenlegen

Ethischer Aspekt: Automatische Datenverarbeitung

- **Dual Use:** Der Einsatz solcher gesammelter Daten für verschiedene Zwecke. Neben dem guten Zweck sie zu Nutzen, um die SuS beim Lernen zu unterstützen wäre es auch möglich sie z. B. zur Überwachung oä zu nutzen.
- **Diskriminierungspotential:** z.B. durch die Aufnahme demografischer Daten
- **Fehlprognosen/-diagnosen:** z. B. durch Offline-Lernaktivitäten oder technologieferne Lernende
- **Black Box:** Die Ergebnisse der KI sind nicht nachvollziehbar, da sie nicht ausgegeben werden. Die oben genannten Punkte können Lehrkräfte dadurch unbemerkt beeinflussen.

Ethischer Aspekt: "Black Box"



healthcare-in-europe.com

News • Wie Künstliche Intelligenz ihre Entscheidungen erklären kann

KI-Diagnostik: Abschied von der 'Black Box'

Wenn ein Algorithmus in einer Gewebeprobe einen Tumor ausmacht, verrät er bislang nicht, wie er zu diesem Ergebnis gekommen ist. Das ist wenig vertrauenswürdig.

Screenshot [ki-diagnostik-abschied-black-box](#) via [healthcare-in-europe.com](#)

Ethischer Aspekt: "Black Box"



"Ein bekanntes Beispiel ist ein Algorithmus, der auf Röntgenbildern Tuberkulose diagnostizieren sollte. Er hat sich unter anderem die Ränder von Röntgenbildern angesehen und dann festgestellt, als er sich selbst trainiert hat, dass auf Bildern von mobilen Röntgengeräten häufiger Tuberkulose zu sehen ist. Das liegt aber daran, dass die mobilen Geräte häufiger in Ländern mit einer hohen Tuberkulose-Prävalenz eingesetzt werden. Das ist natürlich ein völlig medizinfremder Faktor, der keine gute klinische Präzision zulässt. So eine Form der Black Box darf es nicht geben. Sie braucht eine Annotation, um die groben Kriterien zu beschreiben, und dann müssen wir den Algorithmus auch verändern können, damit er ein Kriterium nicht mehr einsetzt."

Medizin 5.0 Beratung durch selbstlernende Algorithmen via www.medica.de



Andere Länder kritischer

Nur jeder vierte Deutsche überprüft KI-Ergebnisse

Stand: 02.05.2025 14:01 Uhr

KI-Chatbots werben mit smarten Antworten auf Knopfdruck. Doch nicht immer sind die Angaben richtig. Dennoch überprüft hierzulande offenbar nur etwa jeder Vierte die Ergebnisse.

Nur gut ein Viertel der Nutzer in Deutschland überprüft einer Umfrage zufolge die Ergebnisse, die KI-Chatbots wie ChatGPT, Google Gemini und Microsoft Copilot erstellen. Konkret gaben 27 Prozent der Befragten hierzulande an, KI-generierte Texte, Bilder oder Übersetzungen gegenzuprüfen. Das zeigt die [internationale Analyse](#) der Prüfungs- und Beratungsgesellschaft EY zum Thema Künstliche Intelligenz (KI). Der Durchschnitt in den 15 untersuchten Länder lag bei 31 Prozent.



Plenumsdiskussion Transparenz & Fehleranfälligkeit



Plenum



15 min

- Welchen Anspruch stellen wir an (guten) Unterricht?
- Sollte Erklärbarkeit ein pädagogisches Qualitätskriterium für KI-Systeme sein?
- Wann wart ihr bereit, einer KI-Entscheidung zu vertrauen obwohl ihr nicht wusstet, wie sie zustande kam?
- Ist ein korrektes Ergebnis ausreichend, auch wenn der Entscheidungsweg intransparent bleibt?
- Verstärkt KI bestehende Bildungsungleichheiten – oder kann sie sie reduzieren?
- In welchen Bereichen des Bildungssystems wären Black-Box-Systeme akzeptabel – und wo nicht?
- Sollte das Bildungssystem bewusst auf etwas Genauigkeit verzichten, um Erklärbarkeit zu erhalten?
- Ist ein falscher KI-Output problematischer als ein menschlicher Fehler – warum oder warum nicht?

Abschluss & Ausblick



Portfolioauftrag: ethische Reflexion



Welche ethischen Aspekte, die in der heutigen Sitzung aufgeworfen wurden, waren *neu für dich* und welche hältst du für deinen Alltag und dein späteres Berufsleben für *besonders relevant und warum?*
(**Nenne** und **Bewerte**)

- **Erkläre** was wir als *Fach- (und KI-)Expert*innen* leisten müssen, um sicherzugehen, dass *KI auf sinnvolle Weise* verwendet wird
- **Beschreibe** wie sich das *konkret im Schulalltag* umsetzen lassen könnte

Sitzungsziele ... erreicht? Zeit für eure Rückfragen



Am Ende dieser Sitzung können die Studierenden...

- Beispiele für *ethische Konflikte* der Verwendung von KI (im Unterricht) **nennen**
- Den *Interpretability-Accuracy Trade-off* **beschreiben**
- Die Begriffe *Kausalität* und *Korrelation* anhand eines Beispiels **erklären**
- Die *Funktionsweise mathematischer Logik mit probabilistischer KI* **vergleichen** und unter verschiedenen Gesichtspunkten (z. B. wissenschaftlich, bildungspolitisch, rechtlich, ...) **diskutieren**

Quellenverzeichnis



- [ki-überzeugt-uns-besser-als-menschen-und-das-sollte-angst-machen](#)
- https://www.medica.de/de/medtech-devices/Medizin_5.0_Beratung_durch_selbstlernende_Algorithmen

Die Präsentation „Von Algorithmen zu KI Teil I“ wurde vom Projektteam „[WoLKE](#)“ erstellt und ist lizenziert unter einer Creative Commons Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International Lizenz [CC BY SA 4.0](#).

Ausgenommen von der Lizenz ist folgendes, zitiertes (Bild)Material:



- Alle in Microsoft PowerPoint enthaltenen Bilder, Symbole, Piktogramme und Designelemente. Hierfür gelten die Nutzungsbedingungen von Microsoft.

Bei der Gestaltung dieser offenen Bildungsressource wurde generative KI (GKI) eingesetzt. Folgende Aufgaben wurden teilweise unter vollständiger menschlicher Aufsicht an GKI-Werkzeuge delegiert:

- Aufbereitung von Inhalten
- Zusammenfassen von Texten
- Bildgenerierung

Folgende(s) GKI-Werkzeug(e) wurde(n) eingesetzt: Dall-E, ChatGPT, Gemini

Die Verantwortung für das fertige Material liegt vollständig bei den Autor*innen.