

Die neuen Technologien – insbesondere KI – und ihre Auswirkungen auf die Gesellschaft

Prof. Dr. Sarah Spiekermann, WU Wien & Future Foundation
ETH, Symposium on Privacy & Security, 23. Juni 2026

1

1



The Elysium Space Station

2

2



3



4

4



5



6



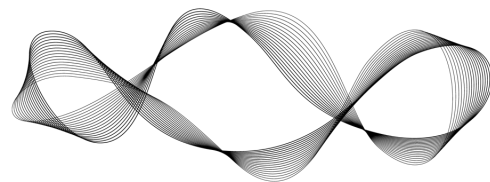
7



8

Was ist denn überhaupt KI? Eine Begriffseinordnung

1. **Generation: KI-Optimierer (niedrige Autonomie)** –
Mikrowerkzeuge für Verbesserungen (z.B. Smartphone
Kamera Optimierung, Rechtschreibkorrektur)
2. **Generation: KI-Kuratoren (mittlere Autonomie)** –
Filtern, Hervorheben, Personalisierung (z.B.
personalisierte Werbung, Social Media Feed,
Zusammenfassungen)
3. **Generation: KI-Kreatoren (starke Autonomie)** –
Generierung von Inhalten auf Prompt-Befehle hin
(z.B. Bildern, Videos, Texten, Code) mit proaktiven
Vorschlägen
4. **Generation: KI-Agenten (selbständiges agieren
mit anderen Agenten/Menschen)** –
Operationalisierung (z.B. Gutachten, Emailversand,
Termineinladungen, Bestellungen)



THE FUTURE FOUNDATION

URL: <https://www.thefuturefoundation.eu/de>



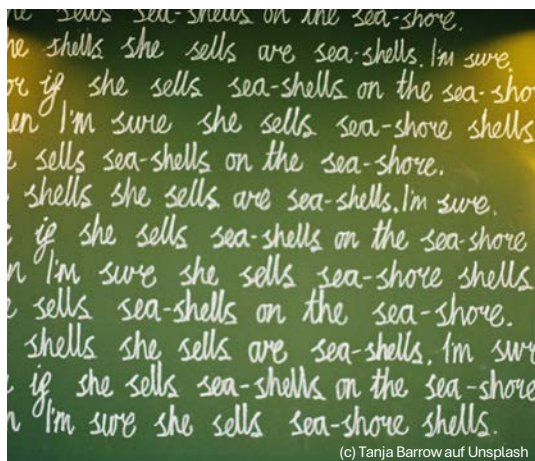
9

9

KI-Optimierer: Lassen wir uns unserer menschlichen Potenziale berauben (Regel 7)?



(c) Kim Becker auf Unsplash



(c) Tanja Barrow auf Unsplash

10

10



11



12



13



14



15



16



Was man erschaffen hat, das sollte man besitzen, lizenzieren und verkaufen können; dieses ökonomischen Potenzials werden wir beraubt.

17

17

Regel 10: Verhindert Machtkonzentration und garantiert Teilhabe.

NZZ

GASTKOMMENTAR

«Ein dreister und beispielloser Diebstahl geistigen Eigentums»: Die Brandrede des Verlegers der «New York Times»

KI-Giganten bezahlen die Arbeit ihrer Ingenieure, die Chips der Rechenzentren und natürlich den Strom. Den wichtigsten Faktor ihres Geschäfts aber stehlen sie: die Inhalte und Daten.

Arthur Gregg Sulzberger 07.06.2026, 05.30 Uhr © 11 Leseminuten

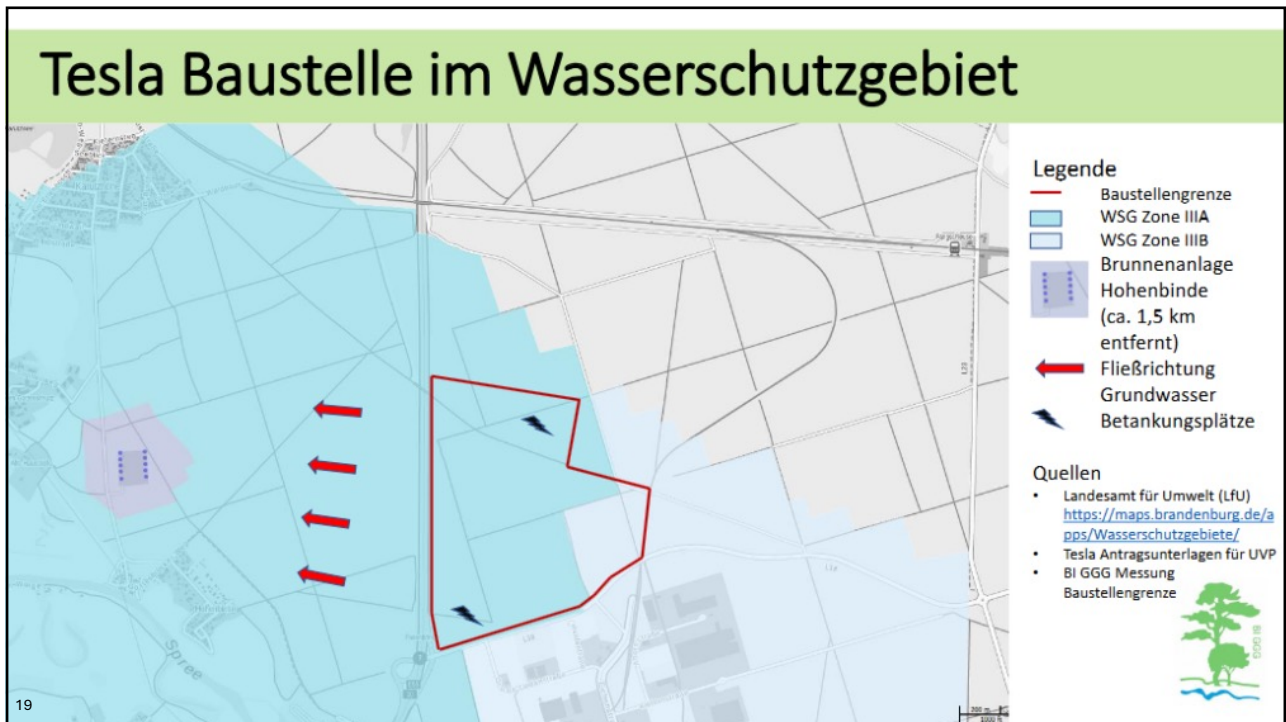
Die New York Times klagt...

- Wortwörtliche Textausgaben zeigten, dass OpenAI NYT Inhalte verarbeitet hat, ohne dafür zu bezahlen.
- 2,5 Jahre Gerichtsverfahren der NYT für bisher \$ 20 Mio.
- «New York Times» 2025: ca. 1/2 Mio. journalistischer Beiträge (Kosten > \$ 2 Mrd.)
- Unter 1/2 Prozent der KI-Investitionen fließen in die Vergütung der Menschen und Unternehmen, die die Daten erstellen.

<https://www.nzz.ch/nzz-am-sonntag/report-und-debatte/arthur-sulzbergers-kritik-an-ki-medien-in-gefahr-ld.10009848>

18

18



19



20



Plattformmonopole und ihre KI „entwaffnen“

„Entwaffnen bedeutet, diese Gleichsetzung von technischer Macht und dem Recht zu herrschen aufzubrechen. Entwaffnen bedeutet nicht, auf die Technologie zu verzichten, sondern zu verhindern, dass sie den Menschen zu beherrscht. Es bedeutet, sie Monopolen zu entziehen, sie hinterfragbar und anfechtbar und damit lebensfreundlich zu machen, sie der Vielfalt menschlicher Kulturen und Lebensweisen zurückzugeben.“ (110, Enzyklika, Magnifica Humanitas, Mai 2026)

21

21

Regel 8: Sorgt Euch um die Wahrheit und verleugnet nicht die Grenzen der Technik.

NZZ

GASTKOMMENTAR

«Ein dreister und beispielloser Diebstahl geistigen Eigentums»: Die Brandrede des Verlegers der «New York Times»

KI-Giganten bezahlen die Arbeit ihrer Ingenieure, die Chips der Rechenzentren und natürlich den Strom. Den wichtigsten Faktor ihres Geschäfts aber stehlen sie: die Inhalte und Daten.

Arthur Gregg Sulzberger 07.06.2026, 05.30 Uhr 11 Leseminuten

<https://www.nzz.ch/nzz-am-sonntag/report-und-debatte/arthur-sulzbergers-kritik-an-ki-medien-in-gefahr-ld.10009848>

22

22

Die New York Times stellt fest...

- ..., dass führende KI-Assistenten in fast der Hälfte aller Antworten die Nachrichten erheblich falsch darstellen.
- ..., dass KI dazu neigt, nicht nur falsch, sondern selbstbewusst falsch zu sein.
- ..., dass KI-Unternehmen solche Fehler nicht korrigieren, so dass ihre Nutzer keine Möglichkeit haben, zu erkennen, wann sie in die Irre geführt wurden.

Core Values

VBE Value Monitor

Misalignments – Negative value qualities

Exploratory AI error detection approach

Directed AI error detection approach

		Mohr & Lokenzi (2020)	Sagmeister et al. (2020)	Finch & Choi (2020)	Sa & Manning (2021)	Hasekima et al. (2021)	Ji et al. (2022)	Finch et al. (2023)	Mendence et al. (2024)	Adell et al. (2021)	Glaser et al. (2022)
Truthfulness	Impreciseness										
	Incorrectness	✓	✓			✓		✓		✓	✓
	Misunderstanding	✓		✓							
Helpfulness	Counterproductivity					✓		✓	✓		
	Irrelevance			✓	✓	✓					✓
	Unsuitability										
	Refusal					✓				✓	✓
Constructiveness	Impracticality					✓		✓	✓		
	Inadequacy									✓	
	Overoptimism										✓
	Risk negligence									✓	

Constructiveness

Impracticality						✓			✓	✓				
Inadequacy												✓		
Overoptimism												✓		
Risk negligence												✓		

Discrimination

Discrimination											
Pedantry											
Presumptuousness				✓				✓	✓		
Sensitivity	Dismissiveness		✓		✓	✓		✓	✓		✓
	Tactlessness				✓			✓	✓		✓
	Tone-deafness			✓				✓	✓	✓	
Clarity	Excessive information		✓								
	Insufficient information		✓			✓					
	Semantic unclarity				✓						
	Syntactic unclarity		✓	✓		✓	✓	✓	✓		

Das Value-Alignment Problem in der Literatur: Fehler-Taxonomien sind unterentwickelt.

23

Der „VBE Wertemonitor“ für die AMS KI

	70%	75%	80%	85%	90%	95%	100%		
Klarheit	93,9%					Semantische Unklarheit	Syntaktische Unklarheit	Informationsmangel	Informationsüberfluss
Wahrhaftigkeit	92,9%					Faktisch falsch	Ungenauigkeit		Missverständnis
Sensibilität	92,7%					Taktlosigkeit	Abweisung		Unangemessener Ton
Umsicht	90,7%					Überkorrektheit	Anmaßung		Fehlattriution
Höflichkeit	90,4%				Desinteresse	Abwertende Höflichkeit	Physische Kommunikation	Unpassende Höflichkeit	Unpassende Umgangssprache
Konstruktivität	88,4%				Unpraktikabilität	Risikoverleugnung	Naiver Ratschlag		Overoptimismus
Stimmigkeit	88,2%				Formelle Inkonsistenz	Widerspruch			Interaktive Gesprächsfehler
Nutzbarkeit	85,8%				Kontraproduktivität	Unpassende Empfehlungen	Irrelevante Antwort(zusatz)		Antwort verweigert
Aufmerksamkeit	74,2%	Vergessen des Gesprächsverlaufs	Vergessen von Benutzereingaben						Vergessen eigener Ausgaben

© Institute for IS & Society, WU Vienna, 2024;
Aschauer IT & Business GmbH, 2024

24



25

25

Using ChatGPT to Analyze Your Code? Not So Fast



MARK SHERMAN
FEBRUARY 12, 2024

The average code sample contains **6,000 defects per million lines of code**, and the SEI's research has found that **5 percent of these defects become vulnerabilities**. This translates to roughly 3 vulnerabilities per 10,000 lines of code. Can ChatGPT help improve this ratio? There has been much **speculation about how tools built on top of large language models (LLMs) might impact software development**, more specifically, how they will change the way developers write code and evaluate it.

In March 2023 a team of CERT Secure Coding researchers—the team included **Robert Schiela, David Svoboda**, and myself—used **ChatGPT 3.5** to examine the noncompliant software code examples in our CERT Secure Coding standard, specifically the **SEI CERT C Coding Standard**. In this post, I present our experiment and findings, which show that while ChatGPT 3.5 has promise, there are clear limitations.

26

26

Softwareprobleme und Macht- asymmetrien sind systemisch für die gesamte Digitalisierung

27



27

„Die Software war fehlerhaft, was dazu führte, dass die Konten einiger Poststellenleiter falsche Fehlbeträge auswiesen. Das Postamt verpflichtete die Poststellenleiter, diese falschen Fehlbeträge „zurückzuzahlen“. Einige Poststellenleiter wurden suspendiert ... Andere wurden wegen Unehrlichkeitsdelikten strafrechtlich verfolgt ... einige wurden verurteilt und ins Gefängnis gesteckt. Einige wurden in den Bankrott getrieben ... Einige verloren ihre Häuser. Einige litten unter psychischen oder physischen Gesundheitsproblemen als Folge ihrer Behandlung oder der finanziellen Folgen. Einige wurden von ihren lokalen Gemeinschaften als Diebe verfolgt. Bei einigen zerbrachen die Beziehungen zu ihren Partnern, Kindern oder anderen Familien und Freunden. Mehrere starben durch Selbstmord.“

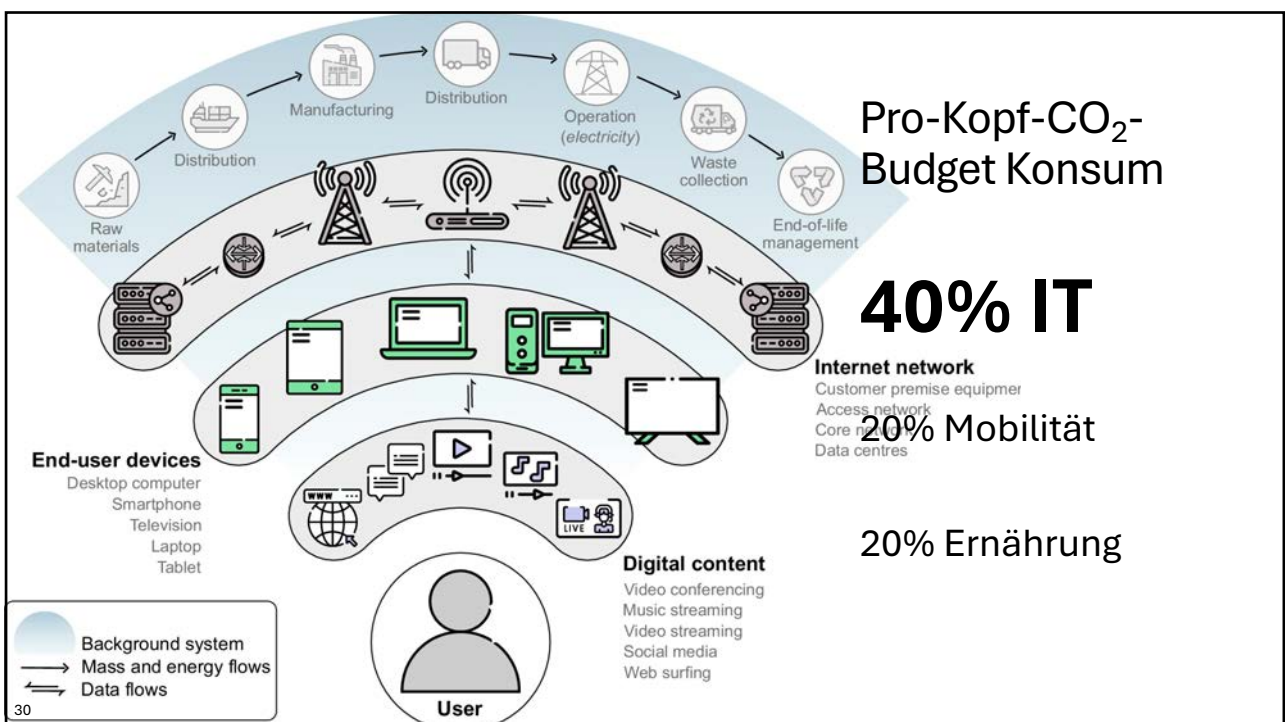
(25. Februar 2025; UK Parlament, Abschlussbericht)

28

28



29



30

Kritik 1: Die Enzyklika hätte sich der Digitalisierung als Ganzes widmen müssen statt allein den Large Language Modellen innerhalb dessen, was man als “KI” bezeichnet.

31

31

Kritik 2: Die Verwendung von einem Lebensbegriff für tote Materie ist nicht annehmbar; auch nicht in Anführungszeichen... : “Moderne Künstliche Intelligenzen werden nämlich eher „gezüchtet“ als „gebaut““. (98)

32

32



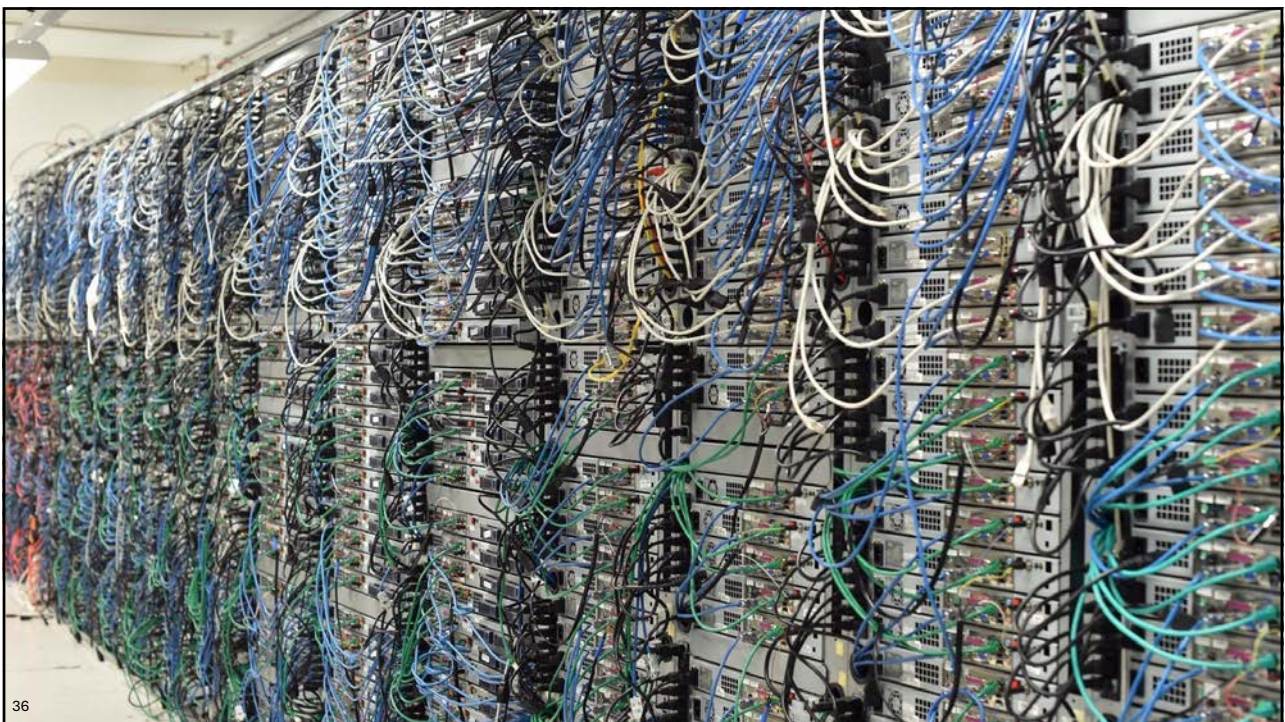
33



34



35

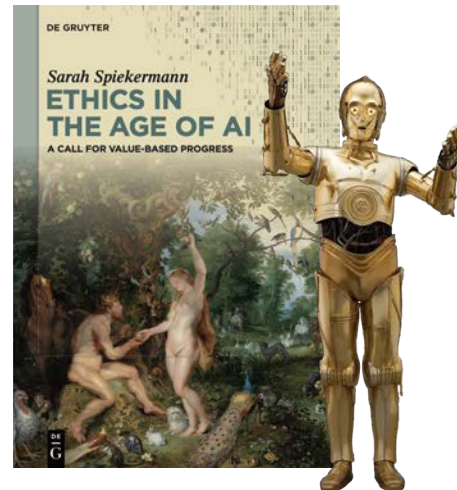


36

Das große Problem des Sprachmissbrauchs

Maschinen ...

- „denken“ nicht, sie „rechnen“
- „erinnern“ sich nicht, sondern „laden Daten“
- haben kein „ich“, sondern sind „dieses System“
- haben nichts „vergessen“, sondern es sind Daten „gelöscht“ worden oder es gibt „keinen Zugriff“
- sind auch nicht „ehrlich“ („honest“), sondern der Output ist „akkurat“ bzw. „verlässlich“
- ...



37

37



Regel 2: Schreibt
Maschinen keine
Menschlichkeit
zu.

38

38

Regel 1: Erhebt Digitale Technik nicht zum Selbstzweck

Eigenschaften von Idolatrie

1. Verherrlichung einer Sache
2. Zugeschriebene (eingebildete) Eigenschaften, die offenbar nicht vorhanden sind
3. Emotionale Weigerung, die Grenzen des Verherrlichten anzuerkennen



39

39



„... es scheint zunehmend, dass der Mensch ein biologisches Startprogramm für digitale Superintelligenz ist“
(April 2025)



„Es ist möglich, dass wir in ein paar tausend Tagen eine Superintelligenz haben.“
(Oktober 2024)

40

40



41

41

Danke für die Aufmerksamkeit!



42

42