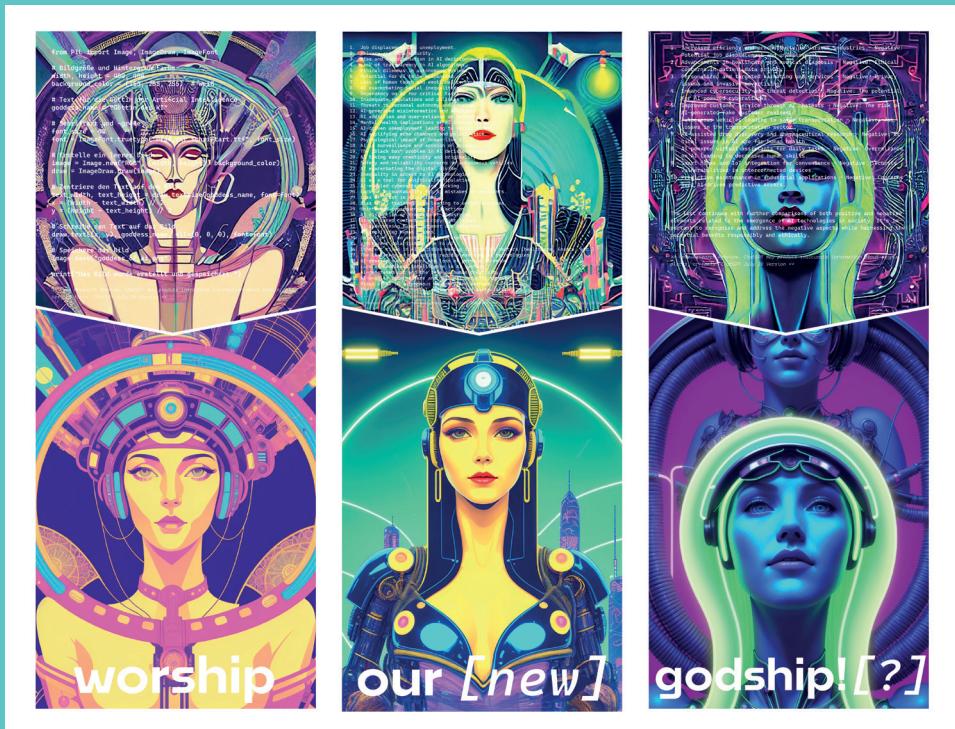


Theo Hug, Petra Missomelius, Heike Ortner (Hg.)

# Künstliche Intelligenz im Diskurs: Interdisziplinäre Perspektiven zur Gegen- wart und Zukunft von KI-Anwendungen





# MEDIEN – WISSEN – BILDUNG

*innsbruck* university press

 universität  
innsbruck



Theo Hug, Petra Missomelius, Heike Ortner (Hg.)

# **Künstliche Intelligenz im Diskurs: Interdisziplinäre Perspektiven zur Gegen- wart und Zukunft von KI-Anwendungen**

Theo Hug

Institut für Medien, Gesellschaft und Kommunikation, Universität Innsbruck

Petra Missomelius

Institut für Medien, Gesellschaft und Kommunikation, Universität Innsbruck

Sprecherin des Forschungszentrums *Innsbruck Media Studies* an der Universität Innsbruck

Heike Ortner

Institut für Germanistik, Universität Innsbruck

Gedruckt mit finanzieller Unterstützung der Moser Holding AG, der Austria Presse Agentur (APA), des Forschungsschwerpunkts *Kulturelle Begegnungen – Kulturelle Konflikte*, des Forschungszentrums *Innsbruck Media Studies* sowie des Vizerektors für Forschung der Universität Innsbruck.

**Tiroler Tageszeitung**

**APA**  
AUSTRIAPRESSEAGENTUR

**universität  
innsbruck**  
Forschungsschwerpunkt Kulturelle  
Begegnungen - Kulturelle Konflikte

 **ims**  
innsbruck  
media  
studies

© *innsbruck university press*, 2024

Universität Innsbruck, Innsbruck – Wien

1. Auflage

Alle Rechte vorbehalten.

[www.uibk.ac.at/iup](http://www.uibk.ac.at/iup)

Titelgrafik: Jaclyn Debiasi; Stable Diffusion (Grafik), Chat GPT (Text)

ISBN 978-3-99106-139-7

DOI 10.15203/99106-139-7

# Inhaltsverzeichnis

Editorial .....	7
<i>Theo Hug, Petra Missomelius und Heike Ortner</i>	
Grußworte zum Medientag 2023 .....	11
<i>Irene Häntsche-Erhart</i>	
KI im Medienalltag – Grußworte zum Medientag 2023 .....	13
<i>Hermann Petz</i>	
Geleitworte zum Medientag 2023 .....	15
<i>Clemens Pig</i>	
Funktionsweise, Möglichkeiten und Grenzen generativer KI .....	17
<i>Justus Piater</i>	
Generative KI in der (Hochschul-)Bildung: Chancen und Risiken .....	29
<i>Doris Weßels &amp; Eike J. Meyer</i>	
Einfluss psychologischer Faktoren auf die KI-Nutzung und -Wahrnehmung .....	43
<i>Christine Anderl, Stefanie H. Klein, Nico Ehrhardt &amp; Sonja Utz</i>	

Künstliche Intelligenz: Menschenähnlich oder Menschzentriert? ..... <i>Maya Pindeus</i>	59
On Coding ..... <i>Miro Roman and Alice_ch3n81</i>	63
Die sprachlichen Fähigkeiten von KI und die Komponente Emotion in der Mensch-Maschine-Interaktion ..... <i>Heike Ortner</i>	79
Zwischen AI Act und Posthumanismus. Künstliche Intelligenz und ihre ethischen „Radialkräfte“ ..... <i>Michael Funk</i>	95
Notizen zum KI-Klima ..... <i>Oliver Leistert</i>	117
Desiderata der AI-Literacy-Diskurse ..... <i>Theo Hug</i>	131
Kurzbiografien der Mitwirkenden .....	151

# Editorial

**Theo Hug, Petra Missomelius, Heike Ortner**

Es ist zu beobachten, wie schnell KI-Anwendungen zu einem selbstverständlichen Teil unserer Gesellschaft, unserer Umwelt und unseres Alltags werden. Sind sie als neue Werkzeuge zu verstehen, die uns helfen, der zunehmenden Komplexität unserer Gesellschaft zu begegnen? Oder verändert der unmerkliche Einsatz dieser Technologien auch uns selbst, unser Menschen- und unser Selbstbild?

Zu den bekanntesten neueren Applikationen der generativen KI zählt zweifellos ChatGPT, ein Chatbot, der auf der Basis von maschinellen Lerntechnologien mit Nutzern und Nutzerinnen über ein Eingabefenster interagiert. Die Antworten, die dabei auf die jeweiligen Eingaben (sog. „prompts“) generiert werden, klingen „natürlich“ und oft plausibel und je nach NutzerInnen-Perspektive auch relevant. Freilich gibt es auch Nonsense und allerhand Einschränkungen, was vielmehr mit dem Label Artificial Stupidity bezeichnet werden sollte.

Die Beiträge dieses Bandes gehen auf den Medientag 2023 sowie der begleitenden Lehrveranstaltung des Forum *Innsbruck Media Studies* der Universität Innsbruck zurück. Sie setzen sich mit Möglichkeiten und Herausforderungen der Nutzung künstlicher Intelligenz im Medienalltag und darüber hinaus auseinander. Dabei beleuchten sie die Frage, was von alltäglichen KI-Anwendungen zu halten ist, wie sie einzusetzen sind und wie sie mit Themen, die unsere Gesellschaft heute, aber vor allem auch in Zukunft beschäftigen werden, zusammenhängen.

Der Medientag sowie die damit verbundene Ringvorlesung mit Übung wurden und werden durch das interfakultäre Forum *Innsbruck Media Studies* in Kooperation mit der Moser Holding AG sowie der Austria Presse Agentur an der Universität Innsbruck veranstaltet. Die Grußworte von Irene Häntsche-Ehart, Vizerektorin für Digitalisierung und Nachhaltigkeit der Universität Innsbruck, Hermann Petz, CEO des Medienunter-

nehmens Moser Holding, und Clemens Pig, CEO der Austria Presse Agentur, leiten zu den Beiträgen in diesem Band über.

**Justus Piater** eröffnet den Band mit einem englischsprachigen Text, in dem er die technisch-konzeptionellen Grundlagen von AI-Systemen wie ChatGPT niederschwellig darlegt. Indem er Möglichkeiten und Grenzen generativer Systeme erläutert, zeigt er auf, warum sie so überzeugende Resultate hervorbringen, man sich jedoch auch der Limitationen dieser Technologien bewusst sein sollte.

Was diese digitaltechnologische Entwicklung für den Bildungssektor bedeutet, beleuchtet der Beitrag von **Doris Weßels** und **Eike J. Meyer**. Sie veranschaulichen anhand der Hochschulbildung, wie wichtig die proaktive Auseinandersetzung mit KI auf allen Ebenen ist, um sie im Bildungsbereich zu regulieren, aber auch zur Förderung und Qualifikation Einzelter nutzen zu können.

Die AutorInnengruppe **Christine Anderl, Stefanie H. Klein, Nico Ehrhardt** und **Sonja Utz** nehmen die neu entstandenen Nutzungsformen von KI-Applikationen unter die Lupe. Anhand neuer Formen der Informationssuche, aber auch der Kreativität, zeigen sie Potentiale zur Steigerung der Kreativität auf, die sie im Umgang mit KI ausmachen. Indem sie die KI-Nutzung und Wahrnehmung untersuchen, liegt ihr Fokus auf dem Einfluss psychologischer Faktoren.

In einem kurzen Impuls wirft **Maya Pindeus** die Frage auf, ob KI menschenähnlich oder nicht vielmehr menschzentriert eingesetzt werden sollte.

Kreativ lotet **Miro Roman** zusammen mit der KI Alice\_ch3n81 im Beitrag 'On Coding' Möglichkeiten im Zusammenleben von KI-Technologie und Mensch aus.

Anknüpfend an die vorherigen Beiträge fokussiert **Heike Ortner** die sprachlichen Fähigkeiten von KI zu emotionaler Kommunikation und der Rolle von Kommunikation in Interaktionen zwischen menschlichen und nicht-menschlichen Entitäten.

Im Zuge der Verbreitung von KI und Fragen der Regulierung ihrer Nutzung geht **Michael Funk** der Frage nach der Relevanz von Ethik nach. Dies untersucht er anhand des AI Act der EU und der Ethik des kritischen Posthumanismus.

**Oliver Leistert** wiederum befragt den Zusammenhang zwischen KI-Technologien und ökologischen Krisen: Ist KI geeignet, der Klimakrise effektiv zu begegnen oder verstärkt sie diese?

Abschließend stellt **Theo Hug** einige Desiderata von AI-Literacy-Diskursen zur Diskussion.

Sowohl die Veranstaltungsorganisation als auch die Herausgabe des Sammelbands waren ein gemeinschaftliches Unternehmen, für das wir in mehrfacher Hinsicht zu danken haben. Unser besonderer Dank gilt unseren beiden Veranstaltungspartnern für ihre Unterstützung, der Moser Holding AG mit Herrn Mag. Hermann Petz und Frau Mag. Silvia Lieb sowie der Austria Presse Agentur mit Herrn Dr. Clemens Pig. Wir danken Frau Mag. Lisa Berger-Rudisch und Frau Barbara Rauch-warter für die kontinuierlich gute Zusammenarbeit. Die Publikation wäre ohne die finanzielle Unterstützung des Vizerektorats für Forschung der Universität Innsbruck, des Forschungsschwerpunkts Kulturelle Begegnungen – Kulturelle Konflikte sowie der beiden Veranstaltungspartner nicht möglich gewesen. Zu danken haben wir außerdem Madeleine Kaiserer für die organisatorische und kommunikative Unterstützung sowie Dr. Birgit Holzner und Carmen Drolshagen von innsbruck university press für die verlegerische Betreuung.

Innsbruck, im August 2024

Theo Hug, Petra Missomelius & Heike Ortner



# Grußworte zum Medientag 2023

**Irene Häntsche-Ehart, Vizerektorin für Digitalisierung und Nachhaltigkeit, Universität Innsbruck**

Liebe Gäste, sehr geehrte Damen und Herren,

ich begrüße Sie im Namen des Rektorats sehr herzlich zum 18. Medientag der Universität Innsbruck!

Der diesjährige Medientag widmet sich dem Thema „KI im Medienalltag“, insofern freut es mich ganz besonders als Vizerektorin für Digitalisierung und Nachhaltigkeit hier begrüßen zu dürfen. Seit dem Launch von ChatGPT durch OpenAI Ende November letzten Jahres ist das Thema KI in aller Munde und in der breiten Öffentlichkeit angekommen. Spätestens seither stellen sich Medien-NutzerInnen die Frage, wer hat den Beitrag verfasst, kann ich mich auf die Information verlassen, wie erkenne ich durch AI generierte Fake-News.

KI zieht aktuell auch in die österreichische Medienlandschaft ein. Harald Kräuter, ORF-Technikchef hat erst kürzlich angekündigt, dass KI-Tools künftig Social-Media-Posts auf Basis von vorhandenem journalistischem Material erstellen sollen.

KI eröffnet zweifellos Potenziale, birgt aber auch Gefahren – wie schon angesprochen. Aber nicht nur im Medien-Kontext ist KI ein Game Changer. Sie stellt uns in allen Lebensbereichen vor Fragen.

Eine umfassende, interdisziplinäre Auseinandersetzung mit diesen, wie sie hier am Innsbrucker Medientag und in der angeschlossenen Ringvorlesung passiert, ist daher unbedingt erforderlich. So hoffe ich, dass die Gespräche und Vorträge dazu beitragen werden, unser aller Verständnis von den Möglichkeiten und Auswirkungen dieser Technologien zu vertiefen und neue Ideen für deren sinnvolle und sichere

Integration in die Medienarbeit, aber auch in den universitären Kontext ganz generell, zu entwickeln.

Mein Dank gilt allen, die zum Gelingen dieser Veranstaltung beitragen bzw. bereits beigetragen haben, insbesondere den Veranstaltern, den Innsbruck Media Studies mit ihrem Sprecher Prof. Hug sowie den langjährigen Kooperationspartnern, der Moser Holding, der APA und dem ORF Tirol. Selbstverständlich gilt mein Dank aber auch allen TeilnehmerInnen für das Interesse und die aktive Beteiligung an diesem wichtigen Diskurs.

Ich wünsche Ihnen allen einen inspirierenden und erkenntnisreichen Medientag!

Vielen Dank!

# KI im Medienalltag – Grußworte zum Medientag 2023

**Hermann Petz, Vorstandsvorsitzender der Moser Holding AG**

Sehr geehrte Frau Vizerektorin Häntsche-Ehart, sehr geehrter Herr Prof. Hug, sehr geehrte Vortragende, sehr geehrte Teilnehmende am Medientag 2023,

wir widmen uns diesmal, wie könnte es im Jahr 2023 anders sein, dem Thema KI im Medienalltag. Bekanntlich beschäftigt sich die Wissenschaft seit vielen Jahren mit dieser facettenreichen Thematik. Seit dem Marktauftritt von ChatGPT im Herbst 2022 hat sich eine Revolution ereignet. KI als Large Language Model auf Basis menschlicher Sprache wurde zum mobilen Alltagswerkzeug, greifbar und universell für alle anwendbar.

Im Alltag von Medienunternehmen spielt KI schon längere Zeit eine immer bedeutender werdende Rolle. Bei der Moser Holding beschäftigt sich die Entwicklungsabteilung seit einigen Jahren mit diesem Thema und tatsächlich wird Künstliche Intelligenz im Journalismus künftig einiges leisten können. Dennoch wird es aus meiner Sicht Bereiche geben, die immer dem Journalisten/der Journalistin als Mensch vorenthalten bleiben werden und – ich denke Sie geben mir hier Recht – auch sollen.

## **Einsatzgebiete von KI im Journalismus**

Sinnvolle Einsatzbereiche von KI in unserem Unternehmen sind beispielsweise Tätigkeiten in klassischer Assistenzfunktion, wie das Kürzen oder das Korrigieren von Texten, um hier ein praktisches Beispiel zu nennen. Auch kann Künstliche Intelligenz dabei helfen, große Mengen von Daten in kurzer Zeit zu analysieren, um Muster und Trends im User-Verhalten auf unseren Webseiten zu erkennen, und damit noch zielgerichteter auf Kunden- bzw. Leserwünsche einzugehen. Weiters erweisen sich KI-Tools als hilfreiche Assistenten für standardisierte Content-Erstellung. Für die

Darstellung von Wetterberichten, Sportergebnissen oder Börsendaten kommt die KI bei verschiedenen Medien bereits erfolgreich zur Anwendung.

In der Regel handelt es sich bei all diesen Tätigkeiten jedoch um Assistenzfunktionen, die Journalistinnen und Journalisten in der Praxis Erleichterungen bringen und ihnen mehr Zeit verschaffen, sich auf anspruchsvollere und auch kreativere Tätigkeiten zu konzentrieren.

## **Grenzen der KI im Journalismus**

So ist im Umkehrschluss davon auszugehen, dass trotz Weiterentwicklung von KI-Technologien es Bereiche geben wird, die dem Journalismus als rein menschliche Leistung vorbehalten bleiben. Etwa, wenn es darum geht, berührende, interessante Geschichten erst einmal zu identifizieren. Journalismus heißt auch, mit offenen Augen seine Umgebung wahrzunehmen, Querverbindungen zu ziehen, Fragen und Themenstellungen zu entwickeln. Also News und Informationen aufzugreifen, die „noch nicht im Internet stehen“. Ein nicht unwesentlicher Punkt betrifft den Investigativ-Journalismus. Wichtige Hintergrundinformationen kommen ans Tageslicht, weil Menschen mit Menschen face to face sprechen und einander vertrauen. Gerade in diesem Zusammenhang spielt der Faktor „Mensch“ eine tragende Rolle.

Check. Re-check. Doublecheck. Diese journalistische Grundregel gehört ebenfalls zum unersetzlichen Handwerk von JournalistInnen. Dass Antworten der KI zumindest derzeit in vielen Fällen noch dazu neigen, zu „halluzinieren“ und dadurch Scheinrealitäten erzeugt werden, ist weitgehend bekannt. Allerdings ist auch klar, dass sich KI Modelle permanent weiterentwickeln und derzeit noch vorhandene Lücken zusehends geschlossen werden.

Aus den vorhin genannten Gründen sehe ich Potenzial für klassische Medien in zweierlei Hinsicht: Zum einen können sie die unterstützende Funktion der KI, wo es hilfreich erscheint, in Anspruch nehmen. Zum anderen können sie als vertrauenswürdige Nachrichtenquellen dafür sorgen, dass KI in ihren Medien nur in einem gewissen, abgegrenzten Raum und mit transparenten Spielregeln zum Einsatz kommt. Gesamtgesellschaftlich wird es an uns allen, insbesondere an den Bereichen der Politik, Wissenschaft und Gesellschaft, liegen, Potenziale der KI im positiven Sinn zu nutzen, aber auch ihre Grenzen klar festzulegen.

# Geleitworte zum Medientag 2023

**Clemens Pig, geschäftsführender Vorstand der APA – Austria Presse Agentur**

Sehr geehrte Frau Vizerektorin Häntscher-Ehart, geschätzter Professor Theo Hug, lieber Hermann Petz, meine sehr verehrten Damen und Herren,

wir schreiben das Frühjahr 2023. Ich mache Bekanntschaft mit ChatGPT und möchte vieles wissen. Unter anderem stelle ich die Frage: „Was ist die APA?“. Eine einfache Abfrage für ein generatives KI-System, könnte man meinen, denn korrekte Beschreibungen und Profile über die APA, ihre Struktur und ihre Eigentumsverhältnisse finden sich in Online- und Offline-Quellen zur Genüge.

Die Antwort von ChatGPT überrascht. Das System weist die staatlich unabhängige Austria Presse Agentur als ein öffentlich-rechtliches Unternehmen aus, das vom Innenministerium finanziert wird. Sie und ich wissen: Das sind echte Fake News und das genaue Gegenteil dessen, was die APA ist, nämlich eine private Agentur im Medieneigentum ohne jegliche staatliche Subvention. Dass diese Art von Falschinformation in Kombination mit starker Verbreitungskraft nicht nur die Reputation von Unternehmen schädigen kann, muss an dieser Stelle nicht erläutert werden.

Selbst wenn uns das genannte Beispiel über die APA oder ein KI-generiertes Bild des Papstes im Designer-Daunenmantel ein kurzes Schmunzeln entlocken mögen, so bleibt uns spätestens bei der aktuellen Flut von Fake-News-Bildern aus der Ukraine oder bei den Desinformationskampagnen, wie sie derzeit im Israel-Palästina-Konflikt kursieren, sehr schnell das Lachen im Halse stecken.

Sehr geehrte Organisatoren und Organisatorinnen des Medientages an der Universität Innsbruck, ich danke Ihnen, dass Sie das Thema Künstliche Intelligenz und ihre Auswirkungen auf Medien und Journalismus und damit auf liberale Demokratien und unsere offene Gesellschaft für ihre diesjährige Veranstaltung aufgegriffen haben.

Der Zeitpunkt könnte nicht besser sein. Wir stehen vor einem Superwahljahr 2024: Gemeinderats- und Landtagswahlen sowie die Nationalratswahl in Österreich, die Wahl zum Europäischen Parlament oder jene zur US-Präsidentschaft. Mit Blick auf diese zentralen Wahlgänge war alles, was wir bisher an KI-generierten Fake News gesehen haben, wohl nur ein Gruß aus der Küche. Als Medien werden wir uns künftig auf immer professionellere, finanziell und personell gut ausgestattete Fake-Akteure mit prall gefüllten KI-Werkzeugkästen einstellen müssen.

Dass wir uns jetzt und in Zukunft geschlossen gegen eine beispiellose Anzahl an Falschnachrichten positionieren müssen, ist nur eine der Auswirkungen, die neue Generationen Künstlicher Intelligenz in unserer digitalen (Des-)Informationsgesellschaft auf uns haben werden. Redaktionen müssen sich noch stärker auf ihre Kernaufgabe konzentrieren und den Fake-Fabrikanten mit faktenbasiertem, ausgewogener und zuverlässiger Information die Stirn bieten.

Gerade weil KI zahlreiche Risiken im Bereich Desinformation und Fake News birgt, gehört sie in das professionelle und zuverlässige Umfeld von Qualitätsmedien. Nur wenn wir die Technologie mit faktenbasierten und zuverlässigen Informationen trainieren, können wir verhindern, dass sie zum Superspreader von Fake News wird und der Qualitätsjournalismus seine Position als Anker für vertrauenswürdige Inhalte stärken und behalten kann. Die wertvollen Archive und tagesaktuellen Produktionen der Medien für billiges Geld an die großen KI-Anbieter zu „verscherbeln“, um deren Sprachmodelle zu trainieren, wäre ein fataler Fehler. Stattdessen müssen die freien und unabhängigen Medien und Nachrichtenagenturen kooperieren und gemeinsame IT-Systeme für das Training von KI schaffen, um über Trusted AI neue redaktionelle und tragfähige Geschäftsmodelle zu erschließen.

Ich sehe hier eine Jahrhundertchance für Medienunternehmen. Die aktuelle Technologie-Revolution ermöglicht es den Medien, ihr bestes und teuerstes Gut, nämlich die Journalistinnen und Journalisten und deren täglichen Output an Texten, Bildern, Grafiken, Videos, Audios und Live-Blogs, durch Kooperation und Kräftebündelung wirtschaftlich wesentlich besser zu schützen.

Ich wiederhole daher auch hier, beim Medientag an der Universität Innsbruck, meine konkrete Vision der Gründung einer europäischen Technologie-Genossenschaft, die den Medien wieder ein Stück weit mehr Autonomie, Vertrauen und Stärke gibt.

In diesem Sinne wünsche ich Ihnen für diesen Medientag spannende Einblicke und Erkenntnisse.

# Generative Artificial Intelligence: How It Works And What It Can And Cannot Do

Justus Piater

## *Abstract*

Recent generative AI systems like ChatGPT produce highly-convincing output in sophisticated language, and exhibit remarkable conversational abilities. This persuasive and human-like behavior naturally raises the user's expectations regarding the truthfulness and reasoning capabilities of such systems. This chapter attempts to provide a foundation for realistic expectations by setting an accessible introduction to the inner workings of such systems, and discusses their limitations.

## Introduction

Artificial Intelligence (AI) has been around for many years. We use it daily without thinking about it, in various specialized incarnations such as the face detector in our smartphone camera or the recommender systems of online retailers and social networks. More recently, the mainstream arrival of generative AI has been drawing enormous attention. For the first time, non-specialists can play with relatively generic AI systems that generate content from text prompts and can engage in sophisticated, almost human-like dialogue. How human-like are these systems in fact? What does it mean when they are called (artificially) *intelligent*? To set a foundation for realistic expectations regarding AI, this chapter provides a brief introduction to functional basics of contemporary AI systems, and explores some consequences regarding their capabilities and limitations.

Artificial Intelligence denotes artificial systems that behave intelligently. What exactly it means to behave intelligently is however difficult to define and has been shifting over the past decades of progress in computer science. One skill set that has always

been understood to require intelligence is playing games. Most AI systems developed for solving perfect-information games largely relied on search techniques, until Google DeepMind presented AlphaGo (Silver et al. 2016), a machine-learning system that spectacularly defeated world-class Go champions. It was followed by AlphaZero (Silver et al. 2018), a more generic machine-learning system able to learn diverse board games at a world-class level, including chess and Go. Since then, *AI* has been largely synonymous with *machine learning*.

In 2020 OpenAI introduced GPT-3, the first Large Language Model (LLM) that made the public headlines. Since its successor ChatGPT took the world by storm beginning in November 2022, popular use of the term *AI* has come to more narrowly denote *generative* AI such as text-to-text and text-to-image systems.

Nevertheless, whatever *AI* is intended to mean, since about 2015 the underlying technology is invariably driven by machine learning. Therefore, to build an intuition for what today's AI systems can and cannot do, it is helpful to obtain a basic understanding of machine learning.

## A Gentle Introduction to Machine Learning

Machine learning denotes artificial systems whose behavior is determined by data and not by algorithms. An algorithm is a step-by-step specification of a method for solving a particular problem. A conventional computer program runs an algorithm to solve a given task directly; a machine-learning system runs algorithms to train and apply a machine-learning model to solve this task. The *same* machine-learning model can be trained to perform *different* tasks by using the *same* algorithms to process different training data.

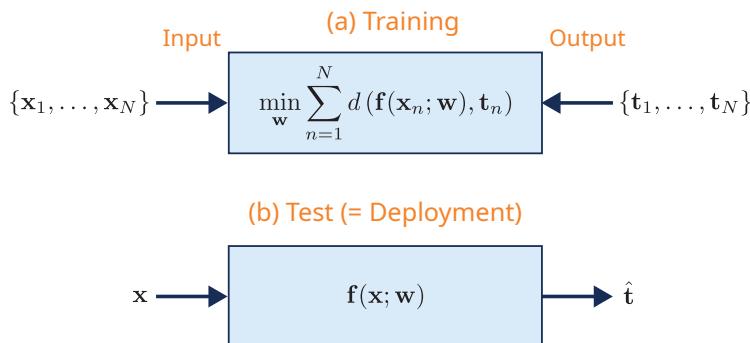


Fig. 1: The supervised machine learning paradigm

Generally, a machine learning model constitutes a function  $\mathbf{f}$  that computes for any input  $\mathbf{x}$  a corresponding output  $\mathbf{t}$ , as illustrated in Fig. 1b. The little hat above the  $\mathbf{t}$  means that this is an *estimate* of the hypothetical, true target value corresponding to  $\mathbf{x}$ . The boldface function and variable names indicate that these variables may be simple numbers or more complex mathematical objects such as vectors, matrices, graphs, etc.

The function  $\mathbf{f}$  is generic; the specific function it computes is determined by its parameter vector  $\mathbf{w}$  (often called a “weight” vector), containing the numbers  $w_1, w_2$ , etc. To get  $\mathbf{f}$  to compute a desired function, we need to adjust these numbers in  $\mathbf{w}$  accordingly. This is done by *training*  $\mathbf{f}$  on a set of  $N$  examples (Fig. 1a), where each example is a pair  $(\mathbf{x}, \mathbf{t})$  consisting of an input  $\mathbf{x}$  and its corresponding, desired output  $\mathbf{t}$  (now without a hat because we know its value; it is not estimated). Mathematically<sup>1</sup>, the objective of training is to find a setting for  $\mathbf{w}$  that minimizes the sum, over the training set, of the distances  $d$  between the predicted and desired outputs. The net effect is that afterwards  $\mathbf{f}$  will reproduce all the  $\mathbf{t}$  from their corresponding  $\mathbf{x}$  as best it can.

Machine-learning models are fitted to training data in the same sense as statistical models are fitted to data. (In fact, the boundary between machine learning and statistics is blurry.) For instance, if the target values  $\mathbf{t}$  are real numbers, then the act of training the model is known as regression. For example, we might train a model to predict the air temperature as a function of the time of day using time series of measured temperatures as training data. If the  $\mathbf{t}$  are categorical, then the trained model performs classification. For example, it might classify images  $\mathbf{x}$  as containing either a cat or a dog.

How well can a machine learning model predict the target values? This depends on its expressive power, which is most of all determined by the number of values contained in its parameter vector  $\mathbf{w}$ . If, say, we limit this number to two and  $\mathbf{f}$  is a polynomial in  $\mathbf{x}$ , then it can only represent straight lines of the form  $f(\mathbf{x}) = w_0 + w_1 \mathbf{x}$ . This is clearly insufficient to represent the oscillation of temperature over multiple days. To avoid this phenomenon of *underfitting* and enable our model to fit any function, we might be tempted to use a very large number of parameters. If we use a model with sufficiently many parameters, we may then be able to train our model to reproduce all our training data points exactly. This is, however, not our ultimate objective; what we really want is that our machine learning model makes good predictions for *unseen* input values; we want it to *generalize*! To this end, fitting the training data exactly is not even desirable.

---

<sup>1</sup> Don’t be scared by the math. This is the most complex mathematical expression of this chapter. The intuition should always be clear from the prose. Mathematical notation is just the best way to nail these concepts precisely and compactly.

For example, outliers – extraordinarily high or low temperature measurements – in the training set will lead to poor predictions under normal circumstances. Disparate temperatures measured at almost the same time of day (on different days or at different locations) will give rise to highly-jagged or even overshooting regression functions, potentially leading to absurd predictions that can be arbitrarily far off the true values. For good generalization, such *overfitting* must absolutely be avoided, most easily in principle by using sufficiently large training sets.

## Neural Networks

Since the earliest days of computing, people have taken inspiration from the biological brain. The first computational model neuron was proposed in 1943 (McCulloch/Pitts 1943), and a mild generalization (Rosenblatt 1958) is still in pervasive use almost unchanged today. The discovery of the training method known as error backpropagation made neural networks practical (Werbos 1975). Due to lacking theoretical foundations (still lacking today) and practical difficulties (which have to some extent been mitigated), interest in neural networks came and went – until in 2012 a neural network, much larger than any earlier networks, won the ImageNet Large Scale Visual Recognition Challenge hands-down, outperforming the competition by a wide margin (Krizhevsky et al. 2012). This set off a flurry of research in neural networks which since then dominate machine learning.

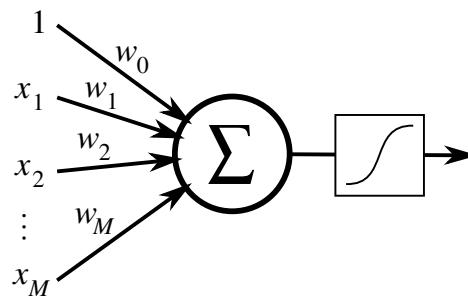


Fig. 2: A model neuron

A model neuron computes a nonlinear function of the weighted sum of its inputs, as shown in Fig. 2. The input includes the numbers  $x_1$  through  $x_M$  as well as the constant 1. These are multiplied by the parameters (weights)  $w_1$  through  $w_M$  and  $w_0$  respectively. These pairwise products are added together, and the result is sent through a nonlinear function (represented by the Greek lowercase letter  $\sigma$ , sigma) that often has an s-shaped (*sigmoid*) shape as illustrated. In summary, this model neuron computes the function

$$f(x_1, x_2, \dots, x_M) = \sigma(w_0 + w_1 x_1 + w_2 x_2 + \dots + w_M x_M).$$

A single model neuron (*unit*) is already quite a capable machine learning model; it can, for example, express linear regression or logistic regression (a probabilistic classification method despite its name).

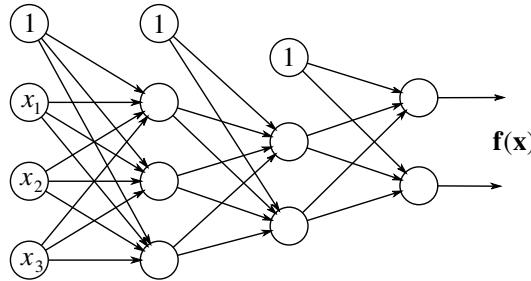


Fig. 3: A small neural network

However, its real power is unleashed if multiple such units are networked together such that the outputs of some units feed into the inputs of downstream units (Fig. 3). The neurons are generally arranged in layers. The more layers, the deeper, we say, the network. Modern neural networks contain tens or even hundreds of layers, which popularized designations such as *deep networks* and *deep learning*. Such networks may contain millions or even billions of parameters, and can represent almost arbitrarily complex functions. On the downside, training them requires enormous amounts of training data and extraordinary computing power to process them.

## Text Processing and Transformers

How can we process natural-language text using neural networks? First, we must encode our input text into a numerical vector that we can feed into the network. To this end, the text is first tokenized, that is, chopped up into a sequence of *tokens* using a fixed algorithm. Each token generally corresponds to a word, to a part of a word, or to a couple of short words that often occur together. These tokens are then converted into vectors such that tokens with related meanings are represented by numerically-similar vectors. Two tokens are deemed related if they tend to occur in similar contexts. The space of these vectors is known as a *word embedding* or *token embedding*. To input a text into a neural network, we can now in principle set the input values of the network to the corresponding values of the embedding vectors representing the tokenized text.

However, this does not allow us to process text of arbitrary length because any given neural network possesses a fixed number of input units. To get around this limitation, the most popular designs for a while used so-called *recurrent* networks that include feedback connections in addition to the feed-forward connections illustrated in Fig. 3, introducing directed loops into the network. The input text is then fed into the network one token at a time. While the input sequence advances, much of the information of already-ingested tokens remains in the network via the feedback connections. Alas, it is very difficult to train recurrent networks to learn relations between specific tokens far apart in the input sequence. Tokens read long ago fade in the network’s memory. Grammatical structure is very difficult to discern in this way.

In 2017 a team at Google published the novel *transformer* architecture for text processing (Vaswani et al. 2017). It contains two key ideas (Fig. 4): First, instead of ingesting tokens sequentially, the embedding vectors of the tokens are augmented by a *position code* that represents the position of the token in the sequence. This allows all tokens of a given input to be fed into the system simultaneously. Secondly, transformed versions of these tokens are correlated with each other in so-called *attention* modules: High correlation between tokens indicates an important relation between them. These transformations are computed by neural networks trained as part of the system. In this way a transformer can learn to pay attention in the sense of which tokens relate to each other under which circumstances. Multiple parallel attention modules, as illustrated in Fig. 4, allows the transformer to consider different types of relations simultaneously. This architecture revolutionized the ability of neural networks to correctly process complex grammatical structure in natural-language text.

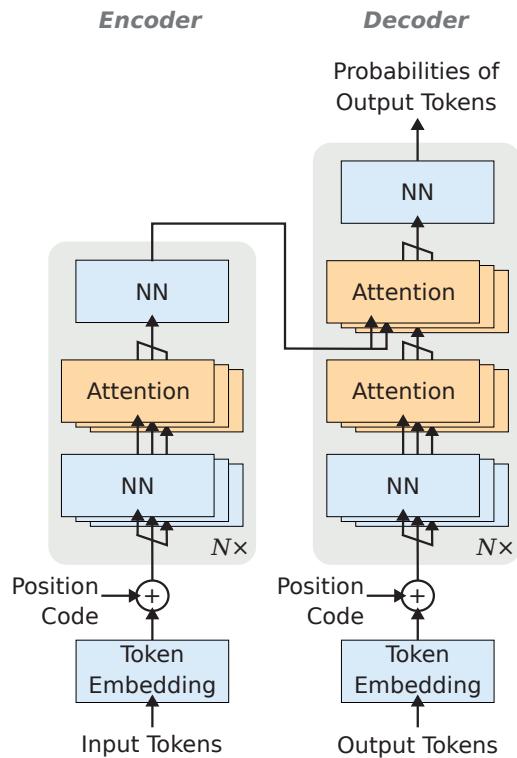


Fig. 4: A simplified overview of the Transformer neural-network architecture

Transformers can be used for diverse text-processing tasks such as classification, sentiment analysis, translation, text generation, dialogue systems, etc. Almost all modern natural-language processing systems are variants of transformers. Because they typically contain millions or billions of parameters they are generally known as Large Language Models (LLM). An LLM is typically *pre-trained*, on large bodies of text, to predict the next word of word sequences. In principle, a training instance is generated by choosing an arbitrary sequence of words from an arbitrary text. The network is trained to predict the last word of the sequence given its predecessors, on millions and millions of such training instances. It can then *generate* new text from input text, one token at a time, by drawing it from the probability distribution output by the system (Fig. 4). The generated text seems natural because the entire partial output sequence is fed back into the transformer, conditioning the probability distribution over the next token to be generated. This pre-training procedure gave rise to the name of the GPT models from OpenAI: Generative Pre-trained Transformer.

## Reinforcement Learning and Alignment

Next-token prediction builds a language model that can continue text but is difficult to get to do anything more goal-directed. To achieve that, the model is further trained to *align* (in OpenAI speak) its answers with the user expectations. This is done using a machine-learning paradigm called *reinforcement learning* (RL). In RL, an agent perceives the *state* of its environment, and chooses an *action* (Fig. 5). Performing this action then affects its environment. At the next time step, the agent perceives the new state of the environment and receives a *reward*. A reward is a simple number that can be positive, negative (punishment) or zero (no reward). Often, non-zero rewards are only issued at the end of a lengthy task. By trial and error, the agent learns a *policy* that maps each state to an action that maximizes the expected sum of future rewards.

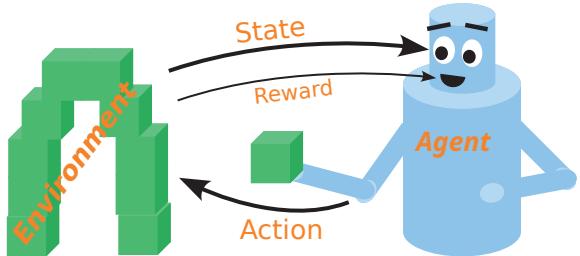


Fig. 5: A high-level view of the reinforcement learning machine-learning paradigm

For training an LLM, the state is a prompt, and the action is the answer generated by the LLM. The reward is given as a function of the quality of the answer. To align the LLM's answers with the expectations of the user, in principle the users have to provide the reward. In practice this reward is first computed from manual rankings of multiple answers by human staff. Since manual ranking is very expensive, a separate machine-learning model is then trained to predict the reward, taking the humans out of the loop and fully automating the RL alignment training.

It is this alignment training (on top of the pre-training) that enables LLM to engage in dialogue, to answer questions, to respond to directives, and to react reasonably and politely to a wide spectrum of fairly-sophisticated queries. The generative process is nevertheless unchanged: Each token from the generated output sequence is drawn from a probability distribution conditioned on its predecessors, on the prompt, and on the history of the current conversation. In particular, this implies that you will typically get different answers each time you ask the same question, and it is quite possible that these answers do not agree and contradict each other.

## The (Limited) Power of LLM

The abilities of LLM have surpassed the expectations of almost everyone, including their developers and most other AI experts. LLM are mere statistical models of language, albeit highly sophisticated and trained on enormous amounts of text. Nevertheless, they are often able to answer remarkably-complex questions correctly, giving the impression of true understanding. On the other hand, they are prone to mistakes, exposing their lack of understanding. This is illustrated by the following example session with ChatGPT 4<sup>2</sup>:

**Me:** Mellie, Peter, Janet and Mike are hanging out in a room. At some point, Peter hands Janet a USB stick and leaves the room for the hallway. Subsequently Mellie leaves the room onto the terrace. Now Janet passes the USB stick on to Mike, who puts it into his pocket. Afterwards Mellie returns into the room, and finally Peter returns. Who does Peter think has the USB stick?

---

<sup>2</sup> This example is inspired by similar scenarios created by my colleague Brigitte Rath to study ChatGPT.

**ChatGPT (1):** [...] when Peter returns to the room, he is most likely to think that Janet still has the USB stick. This is because the last action he witnessed regarding the USB stick was handing it to Janet, and he did not see the exchange between Janet and Mike that happened after he left the room.

This is evidently correct, but this was also fairly easy by GPT-4 standards.

**Me:** After Mellie leaves to go home, Janet tells Peter that she publicly gave the USB stick to Mike while he was out. What does Peter think Mellie thinks who has the USB stick?

Since Peter cannot know that Mellie had also left the room, he will clearly think that she is aware of the USB stick changing hands. But ChatGPT generates an irrelevant line of reasoning and reaches a wrong conclusion:

**ChatGPT (2):** [...] when Mellie left to go home, she was not aware of the conversation that happened between Janet and Peter regarding the USB stick being publicly given to Mike while Peter was out. Therefore, Peter would reason that Mellie thinks Janet still has the USB stick, because Mellie left before this information was shared and the last event she was aware of regarding the USB stick was seeing it with Janet before passing it to Mike, without knowing what Peter was told later.

ChatGPT completely misses the point that Mellie should have witnessed the handover. The verbal communication of this fact is important for Peter but irrelevant, from Peter's perspective, to Mellie's knowledge of the whereabouts of the USB stick. Let's give ChatGPT another chance and generate a new answer:

**ChatGPT (3):** After Janet tells Peter that she gave the USB stick to Mike while Peter was out, and considering Mellie had left to go home [...], Peter would likely think that Mellie believes Mike has the USB stick. This conclusion is based on the premise that Mellie was present in the room when Janet publicly passed the USB stick to Mike. [...]

This time ChatGPT gets it right, including a correct explanation. I follow up with one more question:

**Me:** What does Mellie think who has the USB stick?

**ChatGPT (4):** Since Mellie returned to the room before Peter and was present when Janet publicly gave the USB stick to Mike, Mellie thinks that Mike has the USB stick. [...]

Here, ChatGPT did not pick up that Mellie was in fact absent when Janet gave the USB stick to Mike. Here is another try, yielding a correctly-reasoned answer:

**ChatGPT (5):** Based on the sequence of events and the information provided:

1. Mellie left the room onto the terrace.
2. After Mellie left, Janet passed the USB stick to Mike.
3. Mellie returned to the room, but there's no mention of her being informed about the USB stick's transfer to Mike.

Unless Mellie observed the exchange between Janet and Mike herself or was informed about it before she left to go home, Mellie would likely think that Janet still has the USB stick, because she did not witness the USB stick being passed to Mike and there was no mention of her being told about the transaction afterwards.

In this example session, ChatGPT gets the hard answers sometimes wrong and sometimes right. Now, does ChatGPT really *understand* the questions and its own answers? This is a valid question to ask. In its second answer it follows an irrelevant line of reasoning and misses the real point, and in its fourth answer it gets a fact wrong. On the other hand, its third and fifth answers are argued correctly.

This mixed-bag behavior can be explained by the fact that ChatGPT is mostly trained on just text. There are no mechanisms that explicitly connect the words it sees to their meanings. There is no real-world experience; it has no way to learn what it means to hand an item to somebody else, to observe something, or to leave a room, or even what a room or what an item is. The only connection ChatGPT has to the real world is via the alignment training. While this is only a tiny connection – the only feedback it receives from the real world is numbers representing rewards – I believe it to be quite significant and crucial for its reasoning capabilities. One powerful way to gain new understanding is by using one's current understanding to generate hypotheses and then test them, e.g. by performing experiments or by asking a teacher. The latter happens during alignment training: The LLM produces trial answers and has them evaluated by teachers. This opens a narrow pipe for real-world information to trickle into the LLM.

## Conclusions

It is highly remarkable what levels of quality answers today's LLM can produce. They will undoubtedly be further improved. However, the fundamental problem is unlikely to be solved overnight: Their limited grounding in the real world critically limits their ultimate reliability. No matter how much improvement these systems see, this will remain an issue. Without effectively conveying semantics to the system it will always be possible to ask questions that are not reliably answered correctly, exposing the gap between text statistics and actual meaning. Current research seeks to go beyond text, e.g. by including video. This can be expected to push the boundaries substantially further, since video describes the world in a way that is highly complementary to text. However, video is ultimately just pixels just like text is just letters. By pure observation alone, without interacting with the concepts described by the words, without the ability to formulate and test hypotheses, the ability of machines – and humans – to gain understanding will always be limited.

So, is AI “intelligent”? Without a clear definition of intelligence this question is hard to answer and, unsurprisingly, experts disagree. If we adopt a loose definition of intelligence, it is probably fair to say that smartphone face detectors exhibit some level of intelligence by performing complex pattern recognition, that chess programs exhibit some level of intelligence by solving long-range planning problems in large combinatorial state spaces, and that LLM exhibit some level of intelligence by engaging in sophisticated dialogue. We attribute intelligence to behavior that seems intelligent to us. Absent a rigorous definition of intelligence this is probably not the worst way to use this term. Nevertheless, we must be careful not to overestimate the capabilities and trustworthiness of current AI. It is trained on human-generated data; it does not possess the kind of intelligence required to expand its own capabilities. Current AI lacks many aspects of human intelligence, but it can be very useful in what it can do. Let's use it productively and responsibly.

## References

- Krizhevsky, Alex; Sutskever, Ilya & Hinton, Geoffrey (2012): ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, pp. 1097–1105.
- McCulloch, Warren S. & Pitts, Walter (1943): A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5 (4), pp. 115–133.

- Rosenblatt, Frank (1958): The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 (6), pp. 386–408.
- Silver, David et al. (2016): Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 529, pp. 484–489.
- Silver, David et al. (2018): A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362 (6419), pp. 1140–1144.
- Vaswani, Ashish et al. (2017): Attention is all you need. *Advances in Neural Information Processing Systems*, 30, pp. 5998–6008.
- Werbos, Paul (1975): *Beyond Regression: New tools for prediction and Analysis in the behavioral sciences*. Doctoral Dissertation, Harvard University.

# **Generative KI in der (Hochschul-)Bildung: Chancen und Risiken**

**Doris Weßels, Eike J. Meyer**

## *Zusammenfassung*

Der folgende Beitrag präsentiert eine Analyse der Auswirkungen generativer Künstlicher Intelligenz (KI) auf den Bildungssektor, mit einem besonderen Fokus auf den Hochschulbereich. Der Beitrag ist angelehnt an wissenschaftliche Workshops, in denen basierend auf den vorgestellten Entwicklungen und Perspektiven generativer KI die Potenziale und Herausforderungen, die diese Technologie für Lehr- und Lernprozesse mit sich bringt, diskutiert wurden. Es werden sowohl die bisherige Entstehung als auch bereits heute ersichtliche Auswirkungen beleuchtet. Der Beitrag unterstreicht die Notwendigkeit einer proaktiven Auseinandersetzung mit generativer KI, um die Vorteile auszuschöpfen, gleichzeitig aber auch die Entwicklung von Richtlinien und Rahmenbedingungen zum Management der Risiken und einen (bildungs-)gerechten Zugang zu diesen Technologien zu gewährleisten. Darüber hinaus werden einige der explorativen Ansätze aus den Workshops dargelegt, um einen weiteren schnellen Einstieg in die Materie zu ermöglichen, denn aufgrund der dynamischen Entwicklungen im Themenfeld der generativen KI ist die Qualifikation jedes Einzelnen in Form einer persönlichen, aktiven und kontinuierlichen Auseinandersetzung entscheidend für einen erfolgreichen und zeitgemäßen Einsatz im Hochschulumfeld.

## **Generative KI-Sprachmodelle von OpenAI: Hintergründe, Entstehung, Funktionsweise und aktuelle Entwicklungen**

Künstliche Intelligenz ist bereits seit vielen Jahren ein technologischer Schwerpunkt im Bereich der IT und stellt immer wieder bestehende Prozesse in Organisationen und der Gesellschaft in Frage (vgl. Deckert/Meyer 2020). Doch die Entwicklung generativer Künstlicher Intelligenz (KI), insbesondere im Bereich der Sprachmodelle, markiert einen signifikanten Fortschritt in der Fähigkeit von Maschinen, menschenähnliche Texte

zu analysieren und zu generieren, und gibt damit Maschinen eine neue Art von Zugang zu der Summe digitaler Daten und Informationen. OpenAI, aktuell wohl das führende Unternehmen im Bereich generativer KI, hat mit der Einführung seiner Generative Pre-trained Transformer (GPT)-Reihe maßgeblich zu diesem Fortschritt beigetragen. Die zugrundeliegende Entwicklung begann mit GPT-1 und hat sich bis hin zu den neueren Modellen wie GPT-4o weiterentwickelt. Die Grundidee hinter diesen Modellen ist einfach, doch in ihren Auswirkungen revolutionär: Durch das Training auf umfangreichen Textdaten lernen sie, Zusammenhänge zu erkennen, Vorhersagen zu treffen und darauf basierend neue Texte zu generieren, die in Struktur und Inhalt menschlicher Sprache ähneln (vgl. Limburg et al. 2022).

Die Entstehung der GPT-Modelle basiert auf der Idee des maschinellen Lernens, bei der ein Modell durch die Analyse großer Datenmengen Muster erkennen und lernen kann, ohne explizit programmiert zu werden. Diese Modelle werden zunächst mit einem breiten Spektrum an Textdaten aus dem Internet vorab trainiert, was ihnen ein grundlegendes Verständnis der Sprache verleiht. Anschließend können sie durch weiteres Training bzw. Finetuning auf spezifische Aufgaben oder Wissensgebiete angepasst werden (vgl. Touvron et al. 2023).

Die Funktionsweise von GPT basiert auf dem Transformer-Architekturmodell, das es ermöglicht, die Bedeutung von Wörtern im Kontext zu anderen Wörtern in einem Satz oder Textabschnitt zu verstehen. Dies geschieht durch Mechanismen, die es dem Modell ermöglichen, Verbindungen auch zwischen weit auseinander liegenden Wörtern herzustellen und so die Bedeutung eines Textes präziser zu erfassen.

Die neuesten Entwicklungen in der GPT-Reihe, insbesondere GPT-4, haben diese Modelle noch leistungsfähiger gemacht. Sie können nun nicht nur Texte in menschenähnlicher Qualität generieren, sondern verstehen und produzieren Inhalte in einer Weise, die komplexes Denken und teilweise auch Kreativität imitiert. Dies hat weitreichende Implikationen für zahlreiche Anwendungsbereiche, von der automatisierten Erstellung von Inhalten über die Unterstützung bei der Texterstellung bis hin zur Simulation von Dialogen und Lehrmaterialien.

Trotz dieser Fortschritte bringen die Entwicklungen auch Herausforderungen mit sich, insbesondere im Hinblick auf ethische Überlegungen, Bias in Trainingsdaten und die potenzielle Verwendung der Technologie zur Erzeugung irreführender oder manipulativer Inhalte. Die Forschung und Entwicklung bei OpenAI und anderen Institutionen konzentrieren sich daher nicht nur auf die Verbesserung der Leistungsfähigkeit dieser Modelle, sondern auch auf die Schaffung von Richtlinien und Technologien, um ihre

verantwortungsvolle Nutzung zu gewährleisten. Eine besondere Bedeutung kommt in diesem Kontext auch dem Phänomen der Halluzinationen sowie Filtermechanismen zu, worauf nachfolgend vertiefter eingegangen werden soll.

## Halluzinationen & Faktentreue

Im Kontext von künstlicher Intelligenz (KI) steht der Begriff der Halluzination für das Phänomen, bei dem maschinelle Lernmodelle oder KI-Systeme irreführende und objektiv falsche Informationen generieren. Solche Halluzinationen können in verschiedenen Anwendungen der KI auftreten, wie beispielsweise in der Bilderkennung, der Sprachverarbeitung oder der Textgenerierung. Sie entstehen oft durch Übergeneralisierung, unzureichende Trainingsdaten oder Verzerrungen innerhalb des Datensatzes.

Halluzinationen im Kontext von KI stellen eine besondere Gefahr beim Einsatz von künstlicher Intelligenz in der Hochschulbildung dar, wo die Genauigkeit und Verlässlichkeit von Informationen von größter Bedeutung sind, siehe hierzu die Entwicklung der Faktentreue bei den GPT-Modellen von OpenAI in Abbildung 1.

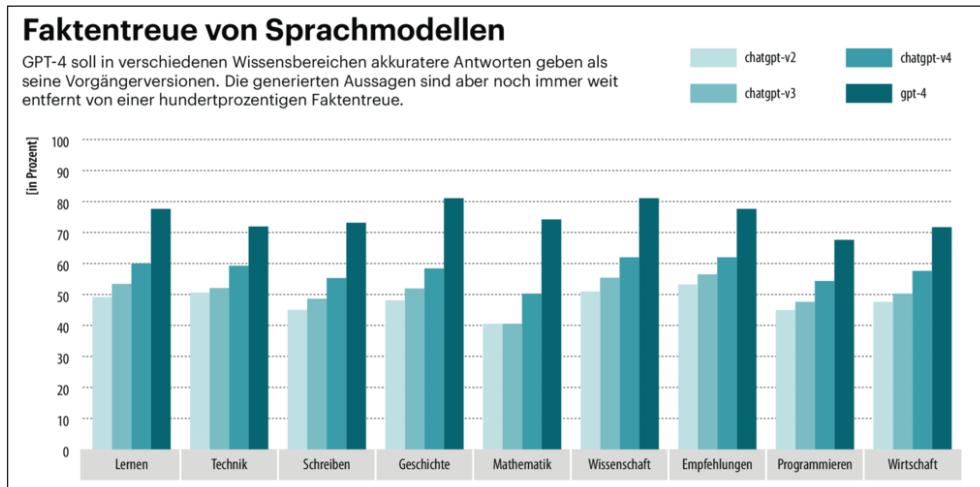


Abbildung 1: Faktentreue bei Sprachmodellen (Schönthaler, 2023)

Beim Einsatz von KI-gestützten Lehr- und Lernwerkzeugen, wie etwa in der automatisierten Generierung von Studienmaterialien, der Beantwortung studentischer Anfragen oder der Bewertung von Aufgaben, können Halluzinationen zu fehlerhaften oder irreführenden Informationen führen. Dies könnte das Lernergebnis negativ beeinflussen,

Missverständnisse fördern und das Vertrauen in den Bildungsanbieter bzw. die digitalen Bildungswerzeuge untergraben.

## **Filtermechanismen**

Die Implementierung von Filtermechanismen bei Input und Output in generativen KI-Werkzeugen wie ChatGPT ist ein zentrales Element bei dem Versuch der Sicherstellung einer ethisch verantwortlichen und rechtskonformen Nutzung dieser Technologien. Diese Filter sind darauf ausgelegt, unangemessene, irreführende oder schädliche Inhalte zu erkennen und zu blockieren, bevor sie vom Modell generiert oder als Eingabe akzeptiert werden. Sie bedienen sich dabei spezieller Algorithmen des maschinellen Lernens, um Muster und Kontexte zu analysieren, die auf potenziell problematische Inhalte hinweisen könnten. Dadurch sollen die Integrität der Interaktionen mit der KI gewahrt und die Verbreitung von Falschinformationen, Hassrede oder urheberrechtlich geschützten Materialien verhindert werden. Zudem unterstützen diese Mechanismen die Einhaltung gesetzlicher Vorgaben und ethischer Standards, indem sie eine Filterung basierend auf geografischen, kulturellen und individuellen Präferenzen ermöglichen.

Die Einführung von Filtermechanismen bei generativen KI-Werkzeugen birgt jedoch auch Probleme. Eines der Hauptprobleme ist die Balance zwischen effektiver Inhaltsmoderation und der Wahrung der Meinungsfreiheit. Übermäßig restriktive Filter können zur ungewollten Zensur von legitimen Inhalten führen und die kreative und informative Nutzung von KI-Technologien einschränken. Ein weiteres Problem ist die Genauigkeit der Filterung; falsch positive oder falsch negative Erkennungen können entweder zu einer ungerechtfertigten Blockierung von Inhalten oder zur Freigabe schädlicher Inhalte führen. Darüber hinaus stellen die kontinuierliche Anpassung der Filter an sich wandelnde soziale Normen und die Vielfalt der menschlichen Sprache eine erhebliche Herausforderung dar. Als Beispiel ist hier die Debatte um die Verzerrung historischer Visualisierungen durch Filter zur Wahrung von kultureller Vielfalt zu nennen (vgl. Raghavan 2024). Schließlich können Datenschutzbedenken auftreten, da die Überwachung und Analyse von Inhalten oft eine umfangreiche Datenverarbeitung erfordert, was die Privatsphäre der Nutzer beeinträchtigen kann.

## **Aktuelle Entwicklungen**

Insgesamt bieten die generativen KI-Sprachmodelle eine faszinierende Perspektive auf das Leistungspotenzial von KI und eröffnen neue Möglichkeiten für die Zukunft der

Mensch-Maschine-Interaktion. Die multimodalen Nutzungsmöglichkeiten und Einsatzgebiete generativer KI nehmen kontinuierlich zu. Während zunächst die Textgenerierung mit den KI-Sprachmodellen im Fokus stand, wurden nach der Bildgenerierung sehr schnell Werkzeuge zur Videogenerierung vorgestellt. Beispielhaft sei Sora von OpenAI genannt. Erste Videos zeigen bereits ein beträchtliches Potenzial, schon mit einfachen Prompts visuell ansprechende Inhalte zu erzeugen (vgl. OpenAI 2024). Diese Entwicklungen und die damit verbundenen Auswirkungen auf den Bildungssektor – und auch für andere Sektoren in der Wirtschaft und Gesellschaft – sind nur schwer abschätzbar. Es besteht jedoch kein Zweifel daran, dass sie in den kommenden Jahren ein zentraler Innovationsmotor sein werden. Eine nähere Betrachtung aktueller und zukünftiger Anwendungsfälle soll im folgenden Kapitel erfolgen.

## **Einsatz und Bewertung generativer KI im Bildungssektor**

Einer der Hauptvorteile generativer KI im Bildungsbereich ist die Fähigkeit, maßgeschneiderte Lernerfahrungen und individualisierte Lernpfade zu schaffen. Studien zeigen, dass solche personalisierten Lernansätze die Motivation und das Engagement der Lernenden steigern können, was zu besseren Lernergebnissen führt. Durch die Analyse von Daten über Lernstile, Vorwissen und Fortschritt einzelner Lernender können KI-Modelle personalisierte Lernmaterialien und Aufgaben generieren (vgl. Lund/Wang 2023, S. 26–29). Dies trägt dazu bei, den Lernprozess effizienter und effektiver zu gestalten, indem es den individuellen Bedürfnissen der Lernenden entspricht.

Ein weiterer Forschungsschwerpunkt liegt auf der Automatisierung bei der Erstellung von Lehrmaterialien für die Lehrenden. Generative KI kann dabei helfen, zeitaufwendige Aufgaben wie die Erstellung von Lehrtexten, Übungsaufgaben und sogar Prüfungen zu vereinfachen. Lehrkräfte können diese Technologien nutzen, um hochwertige Materialien zu generieren, die auf die spezifischen Lernziele und den Kontext ihres Unterrichts zugeschnitten sind (vgl. Haverkamp 2023). Dies ermöglicht es den Lehrkräften, sich mehr auf die pädagogische Interaktion und individuelle Förderung ihrer Schülerinnen und Schüler zu konzentrieren.

## **Qualität generierter Inhalte und Datenschutz**

Trotz der positiven Aussichten werfen aktuelle Studien auch kritische Fragen auf, insbesondere im Hinblick auf die Qualität der von KI generierten Inhalte und die ethischen Aspekte ihres Einsatzes (vgl. Gill et al. 2024, S. 21-23). Die Bewertung der Genauigkeit

und Angemessenheit dieser Materialien ist entscheidend, da Fehler oder irreführende Informationen negative Auswirkungen auf den Lernprozess haben können. Zudem gibt es Bedenken hinsichtlich des Datenschutzes und der potenziellen Verzerrungen (Bias) in KI-generierten Inhalten, die bestehende Ungleichheiten im Bildungssystem verstärken könnten. Auch wenn die Häufigkeit und Auswirkungen von Halluzinationen, Filtern und anderen Fehlern in den ausgegebenen Informationen durch generative KI kontinuierlich reduziert werden, ist eine kritische Auseinandersetzung mit den Inhalten weiterhin notwendig.

## **Einsatz von KI durch Studierende im akademischen Schreiben**

Auch beim Einsatz durch Studierende gibt es sowohl positive wie auch negative Auswirkungen. Bei einem Einsatz von ChatGPT war so zum Beispiel eine Steigerung der Qualität bei Schreibaufgaben bei gleichzeitiger Senkung des Zeiteinsatzes messbar (vgl. Noy/Zhang 2023). Die Dualität des Technologieeinsatzes erfordert es, auch die potenziellen negativen Auswirkungen des KI-Einsatzes zu betrachten. Dazu gehören die Risiken einer übermäßigen Abhängigkeit von KI-gestützten Werkzeugen, die mögliche Schwächung der Entwicklung eigenständiger Schreibfähigkeiten und des kritischen Denkens sowie Fragen der akademischen Integrität. Die Herausforderung besteht darin, die optimale Balance zu finden, bei der Technologien wie ChatGPT als Ergänzung zur traditionellen Bildung eingesetzt werden, um die Lernergebnisse zu verbessern, ohne die fundamentalen Bildungsziele zu kompromittieren (vgl. Buck/Limburg 2023).

## **Die Rolle der Lehrkraft**

Der (disruptive) Wandel der eingesetzten Bildungstechnologien führt auch zu einem Rollenwandel für die Lehrenden. Aktuelle Studien zeigen die Notwendigkeit einer kritischen Auseinandersetzung mit der Rolle von Lehrkräften in einer zunehmend von KI unterstützten Bildungslandschaft (vgl. Çelik et al. 2022) auf. Während KI große Chancen bietet, die Lehr- und Lernprozesse zu verbessern, betonen Forschungsarbeiten auch den Wert und die Unverzichtbarkeit menschlicher Lehrkräfte, besonders in Bezug auf die Förderung kritischen Denkens, sozialer Interaktion und emotionaler Unterstützung der Lernenden. Diese Relevanz haben BildungsexpertInnen bestätigt, die sich im August 2023 im Rahmen des „Strategy Labs“ der Digital Society Initiative (DSI) unter Leitung der Universität Zürich mit den längerfristigen Auswirkungen von KI an den Hochschulen auseinandergesetzt haben. Das im August 2024 veröffentlichte Positionspapier betont den Wert dieser Kompetenzen, den Lehrende unbedingt vermitteln müssten, um

den Studierenden eine reflektierte und verantwortungsvolle Nutzung von KI-Tools zu gewährleisten (vgl. Bernstein et al. 2024):

- Soziale Kompetenzen: Um der Gefahr entgegenzuwirken, dass der Einsatz von KI-Technologien zu einer Isolation der Studierenden führt, ist es entscheidend, soziale Lernprozesse, Empathie, Widerstandsfähigkeit sowie die Fähigkeit zur effektiven Zusammenarbeit in Teams zu stärken.
- Kritisches Denken: Damit KI-Tools nicht unreflektiert eingesetzt werden, sollten Studierende dazu befähigt werden, kritisch zu hinterfragen, Modelle und Abstraktionen zu durchdenken und unterschiedliche Perspektiven in ihre Analysen einzubeziehen.
- Handlungskompetenz unter Unsicherheit: Um mit der raschen technologischen Entwicklung und globalen Herausforderungen wie dem Klimawandel umgehen zu können, ist es wichtig, Studierenden Fähigkeiten zu vermitteln, die es ihnen ermöglichen, auch unter Unsicherheit fundierte Entscheidungen zu treffen.

## Anforderungen von Studierenden

Analog zu den zuvor erläuterten Veränderungen für die Lehrkräfte, ist auch die Perspektive der Studierenden ein zentraler Faktor bei der Auseinandersetzung mit generativer KI und deren Einsatz. Seit Mai 2021 wurden in jedem Semester an der Kieler Fachhochschule KI-Schreibwerkstätten im Rahmen der interdisziplinären Projektwochen durchgeführt. Die KI-Schreibwerkstatt im Mai 2023 konnte erstmalig auch ChatGPT, veröffentlicht am 30.11.2022, als Tool nutzen. Wie bei allen anderen KI-Schreibwerkstätten zuvor wurde das Ziel verfolgt, den Studierenden ein fundiertes Verständnis für KI-Textgeneratoren zu vermitteln, deren praktische Anwendung zu erproben und den Einsatz dieser Tools kritisch zu reflektieren.

Im Rahmen dieses dreistündigen Online-Workshops, der besonders auf praxisorientierte Phasen fokussiert war, nahmen 27 Studierende aus verschiedenen Fachrichtungen der FH Kiel teil. Während des Workshops standen den Teilnehmenden die KI-basierten Werkzeuge ChatGPT und fobizz zur Verfügung, wobei es ihnen freigestellt war, auch andere KI-gestützte Schreibprogramme oder Textgeneratoren zu nutzen.

Den Abschluss des Workshops bildete eine Diskussionsrunde mit Reflexionsfragen. Dabei wurden Fragen zur Motivationsförderung beim KI-gestützten Schreiben, zu den größten Herausforderungen sowie Empfehlungen für andere Studierende diskutiert. Besonders aufschlussreich waren die Antworten auf die Frage nach den Erwartungen

an die Lehrenden im Umgang mit KI-Tools. Die Antworten in Tabelle 1 spiegeln die „Wünsche an die Lehrenden“ beim Umgang mit Tools und Werkzeugen der generativen KI wider. Sehr deutlich sichtbar ist der Veränderungsdruck, der durch Werkzeuge wie ChatGPT im Hochschulkontext entstanden ist.

<b>Integration und Anwendung in der Lehre</b>
<ul style="list-style-type: none"><li>• Forderung einer offenen Haltung der Lehrenden gegenüber KI-Tools.</li><li>• Förderung der kritischen Auseinandersetzung mit technischen und ethischen Aspekten der KI.</li><li>• Suche nach einem Mittelweg, der die Nutzung von KI-Tools anerkennt und zugleich die Wichtigkeit eigener Gedanken und Kreativität betont.</li><li>• Bewusster Umgang mit KI-Tools unter Berücksichtigung der aktuellen Grenzen und Anpassung an spezifische Fachbereiche.</li><li>• Akzeptanz von KI-Tools als Teil der modernen Arbeitswelt.</li><li>• Wunsch nach einheitlichen Rahmenbedingungen und Klarheit zur Nutzung.</li></ul>

*Tabelle 1: Studentische Wünsche an die Lehrenden im Umgang mit generativer KI*

Die Akzeptanz von KI als festem Bestandteil der modernen Arbeitswelt war für die Studierenden ein zentrales Anliegen, begleitet von dem Wunsch nach klaren, einheitlichen Rahmenbedingungen zur Nutzung dieser Tools im akademischen Kontext. Insgesamt verdeutlicht die Reflexion der Studierenden, dass eine Auseinandersetzung mit KI-Tools nicht nur eine technologische Frage darstellt, sondern auch eine didaktische und ethische Herausforderung für die Lehre, die in Zukunft stärker in den Fokus rücken sollte.

## **Empfehlungen für den explorativen Aufbau von Kompetenzen im Umgang mit KI**

Die Geschwindigkeit und Dynamik, mit der sich insbesondere generative KI weiterentwickelt, erfordern eine kontinuierliche Adaption und Weiterbildung von Verantwortlichen in entsprechenden Anwendungsfeldern. Da die Veröffentlichung technischer Neuerungen hier oft dem aktuellen Forschungs- und Publikationsstand zur praktischen Anwendung von KI voraus ist, ist eine aktive und explorative Herangehensweise unerlässlich (vgl. Meyer/Weßels 2023). Eine abwartende Haltung in der Aneignung von Wissen ist aufgrund der Tragweite der Auswirkungen generativer KI im akademischen Bereich keine Option. Um diesem Umstand Rechnung zu tragen, wurden bereits seit dem Aufkommen der ersten Versionen von ChatGPT immer wieder Workshops durchge-

führt. Sie hatten das Ziel, den Umgang mit entsprechenden Technologien und auch einer sinnvollen praktischen Anwendung zu pilotieren und Best Practices zu identifizieren sowie auch mögliche Vorgehensweisen für den Umgang mit zukünftigen Entwicklungen abzuleiten. Im Folgenden werden Herangehensweisen und Ansätze dargestellt, welche im Rahmen dieser Workshops identifiziert wurden. Diese Erkenntnisse sollen die Entwicklung und Förderung der kritischen Digitalkompetenz der Lernenden im (Hoch-)Schulumfeld unterstützen.

Es gibt viele Möglichkeiten des praktischen und niedrigschwelligen Experimentierens mit generativer KI. ChatGPT von OpenAI, welches aufgrund seiner Leistungsstärke ein offensichtlicher Kandidat für erste Schritte ist, war in der 4o Version kostenfrei zugänglich. Aufgrund von Unterschieden gerade in der Qualität und dem Umfang der Nutzung ist je nach Anbieter und Modell jedoch möglicherweise die Nutzung eines kostenpflichtigen Zugangs sinnvoll. Viele Bildungseinrichtungen haben unterschiedliche Nutzungsmöglichkeiten für KI-Sprachmodelle für die Lehrenden und Lernenden geschaffen. Die Herausforderung bestand darin, den Anforderungen des Datenschutzes (hier der DSGVO) gerecht zu werden und das finanzielle Budget für die Lizenzen oder Nutzungsentgelte bereitzustellen.

Daneben gibt es auch diverse andere Systeme und aufgrund des kontinuierlichen technischen Fortschritts ist eine Produktempfehlung hier nur als temporär zu betrachten. Anbieter offener Standards und Systeme wie Mistral oder auch globale Wettbewerber wie Google mit seinem System Gemini und Meta mit Llama sind weitere Kandidaten.

Nach einer ersten Einarbeitung in die grundlegende Bedienung entsprechender Plattformen sei an dieser Stelle die anwendungsbezogene Exploration nahegelegt, um die Funktionsweise in einem persönlich relevanten Kontext zu erfahren. Diese Herangehensweise vereinfacht die weiteren Lernschritte und ermöglicht darüber hinaus die schnelle Schaffung eines echten Mehrwerts. Aktuelle Beispiele aus Workshops sind angepasste Lehrbots, welche mit eigenen Lehrmaterialen trainiert werden können. Diese können dann im Anschluss teilweise auch direkt unter Einbeziehung Lernender pilotiert werden, um schnell relevantes Feedback zu erhalten.

Wichtig ist es, dieses Vorgehen als erste Schritte auf dem Weg einer kontinuierlichen Auseinandersetzung mit generativen KI-Systemen zu begreifen und nicht als finale Lösung, denn die schnelle technische Entwicklung lässt vermuten, dass es sich hier erst um den Anfang einer umfassenden (R)Evolution des Bildungssektors handelt.

## **Aktueller Trend: Lehrbots / Lernbuddies**

Der Einsatz von generativer Künstlicher Intelligenz (KI) in der akademischen Bildung, insbesondere durch die Entwicklung von Lernbots durch Universitäten, markiert einen spannenden Wendepunkt in der Art und Weise, wie Bildungseinrichtungen Lehr- und Lernprozesse gestalten. Diese Technologie bietet das Potenzial, die Bildungslandschaft grundlegend zu transformieren, indem personalisierte, skalierbare und jederzeit verfügbare Lernangebote bereitgestellt werden können.

Zum einen ermöglichen es Lernbots, die Lerninhalte individuell auf den Wissensstand, die Lerngeschwindigkeit und die Vorlieben jedes Studierenden anzupassen, was eine persönlichere und effizientere Lernerfahrung schafft. Die Zugänglichkeit von Bildung verbessert sich ebenfalls, da KI-gestützte Lernplattformen Studierenden rund um die Uhr zur Verfügung stehen können. Zum anderen bieten sie die Möglichkeit, hochwertige Bildungsangebote zu skalieren, ohne dass im gleichen Maße zusätzliche Lehrkräfte benötigt werden. Dies ist besonders in Fachgebieten von Vorteil, in denen es an qualifizierten Dozenten mangelt. Darüber hinaus können Lernbots die Lehrenden unterstützen, indem sie Routineaufgaben übernehmen, was den Lehrkräften mehr Zeit für forschungsintensive und kreative Aufgaben gibt.

Trotz der vielversprechenden Vorteile müssen die Risiken und Herausforderungen, die mit dem Einsatz von Lernbots einhergehen, bei ihrem Einsatz berücksichtigt werden. Die Qualität der von Lernbots bereitgestellten Inhalte muss überwacht und bei Problemen auch zeitnah adressiert werden, um sicherzustellen, dass die Informationen korrekt und aktuell sind und kein falsches Wissen aufgebaut wird.

Es ist eine sorgfältige Abwägung und Umsetzung erforderlich, um Qualität, Ethik und Zugänglichkeit zu sichern. Die Zukunft wird hier nach aktuellem Stand in einer sinnvollen Integration von KI in den Bildungsprozess liegen, welche die bestehenden klassischen Angebote ergänzt und erweitert. Aktuell gibt es hier bereits Pilotierungen auch an deutschen Hochschulen, beispielsweise im Kontext von einzelnen Modulen wie des AI4PM Tutors (<https://poe.com/AITutor4PM9>) an der Fachhochschule Kiel oder beim Einsatz modulübergreifender Lernbots wie Syntea an der IU (vgl. Möller et al. 2022).

## Schlusswort

Im Bildungsbereich eröffnen sich durch den Einsatz von KI-gestützten Lernplattformen und -tools Möglichkeiten für eine individualisierte Lernerfahrung, die auf die spezifischen Bedürfnisse und Fähigkeiten jedes Lernenden zugeschnitten ist. Dies kann nicht nur die Effizienz des Lernprozesses erhöhen, sondern auch dazu beitragen, die Motivation der Studierenden zu steigern und Bildungsbarrieren abzubauen. Gleichzeitig birgt die Integration von KI in den Lehr- und Lernprozess Risiken, insbesondere im Hinblick auf Datenschutz, ethische Bedenken und die Qualitätssicherung von Lehrinhalten. Daher erfordert die Implementierung von KI-Technologien in der Hochschulbildung eine sorgfältige Planung, regelmäßige Evaluierung und Anpassung, um sicherzustellen, dass diese Technologien ethisch verantwortungsvoll genutzt werden und einen echten Mehrwert für den Bildungsprozess darstellen.

Aufgrund der rasanten technischen Entwicklungen von KI-Anwendungen und deren Reifegrad in den letzten Jahren ist die Auseinandersetzung mit deren Anwendungsmöglichkeiten in fast allen Bereichen ein zentraler Schritt, um die Vorteile dieser Technologie als Gesellschaft gewinnbringend zu nutzen. Da die Weiterentwicklung im Kontext aktueller Trends wie Lehr-/Lernbots und Multimodalität kontinuierlich voranschreitet, wird auch auf Ebene der einzelnen Lehrpersonen eine proaktive Auseinandersetzung mit diesem Thema empfohlen. Daneben ist dies im Kontext der kontinuierlichen Evolution von KI nicht als einmalige, abschließende Tätigkeit zu begreifen, sondern als iterativer, immer wiederkehrender Prozess.

## Literatur

- Bernstein, Abraham; Cordoba, Anaïc; Witt, Claudia; Siegers, Daniel; Weßels, Doris; Thouvenin, Florent; Hastings, Janna; Hatt, Jean-Michel; Buhmann, Joachim; Moreno, Judit Martinez; Schawinski, Kevin; Christen, Markus; Coors, Michael; Denner, Michael; Hunger, Patrick; von Thiessen, Raphael; Fabrikant, Sara; Friemel, Thomas; Neupert, Titus & Sachs, Uli (2024). *KI in Bildung, Forschung und Innovation*. DSI Strategy Lab 2023 – Positions Papier. Digital Society Initiative (DSI).
- Buck, Isabella & Limburg, Anika (2023): Hochschulbildung vor dem Hintergrund von Natural Language Processing (KI-Schreibtools). In: *die Hochschullehre*, 9/2023, S. 70–84.
- Celik, Ismail; Dindar, Muhterem; Muukkonen, Hanni & Järvelä, Sanna (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66 (4), S. 616–630.

- Deckert, Ronald & Meyer, Eike (2020): *Digitalisierung und Künstliche Intelligenz*. Wiesbaden: Springer Gabler.
- Haverkamp, Hendrik (2023): Prüfen mit digitalen Instrumenten. In: *Pädagogik*, 6/2023, S. 10–14.
- Limburg, Anika; Mundorf, Margret; Salden, Peter & Weßels, Doris (2022): Plagiarismus in Zeiten künstlicher Intelligenz. In: *Zeitschrift für Hochschulentwicklung*, Themenheft Akademische Kultur und Wissenschaftsfreiheit angesichts der Digitalisierung von Lehren und Lernen, 17(3), Oktober 2022, S. 91–106.
- Lund, Brady D. & Wang, Ting (2023): Chatting about ChatGPT: how may AI and GPT impact academia and libraries? in: *Library Hi Tech News*, 40 (3), S. 26–29, [online] doi:10.1108/lhtn-01-2023-0009.
- Gill, Sukhpal Singh; Xu, Minxian; Patros, Panos; Wu, Huaming; Kaur, Rupinder; Kaur, Kamalpreet; Fuller, Stephanie; Singh, Manmeet; Arora, Priyansh; Parlikad, Ajith Kumar; Stankovski, Vlado; Abraham, Ajith; Ghosh, Soumya K.; Lutfiyya, Hanan; Kanhere, Salil S.; Bahsoon, Rami; Rana, Omer; Dustdar, Schahram; Sakellariou, Rizos & Steve Uhlig (2024): Transformative Effects of ChatGPT on Modern Education: Emerging Era of AI Chatbots. [online] [https://www.researchgate.net/publication/371347113\\_Transformative\\_Effects\\_of\\_ChatGPT\\_on\\_Modern\\_Education\\_Emerging\\_Era\\_of\\_AI\\_Chatbots](https://www.researchgate.net/publication/371347113_Transformative_Effects_of_ChatGPT_on_Modern_Education_Emerging_Era_of_AI_Chatbots) [Stand vom 11-09-2024].
- Meyer, Eike & Weßels, Doris (2023): Natural Language Processing im akademischen Schreibprozess – mehr Motivation durch Inspiration? In: Schmohl, T., Watanabe, A.; Schelling, K. (Hrsg.): *Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens* (Hochschulbildung. Lehre und Forschung, Bd. 4). Bielefeld: transcript; S. 227–252.
- Möller, Moritz; Nirmal, Gargi; Fabietti, Dario; Stiersdorfer, Quintus; Zakhvatkin, Mark; Sommerfeldt, Holger & Schütt, Sven (2024): Revolutionising Distance Learning: A Comparative Study of Learning Progress with AI-Driven Tutoring. IU International University of Applied Sciences, Abgerufen unter: <https://iu-international-university-of-applied-sciences-research-papers.s3.eu-central-1.amazonaws.com/pfjknzpkrcffpfddy-ypgztbyggxxwnxw.pdf> [abgerufen am 28.02.2024].
- Noy, Shakked & Zhang, Whitney (2023): Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence, in: *Social Science Research Network*. DOI: 10.2139/ssrn.4375283.
- OpenAI (2024): Sora: Creating video from text, [online] <https://openai.com/sora#research> [abgerufen am 28.02.2024].
- Raghavan, Prabhakar (2024): Gemini image generation got it wrong. We'll do better., in: *Google*, 23.02.2024, Abgerufen unter: <https://blog.google/products/gemini/gemini-image-generation-issue/> [Stand vom 11-09-2024].

Schönthaler, Philipp (2023): Schneller als gedacht – ChatGPT zwischen wirtschaftlicher Effizienz und menschlichem Wunschedenken. in: *c't Heft* 9/2023 S. 126-131, Abgerufen unter: <https://www.heise.de/select/ct/2023/9/2304715244186570960> [Stand vom 11-09-2024].

Touvron, Hugo; Louis Martin; Kevin Stone; Peter Albert; Amjad Almahairi; Yasmine Babaei; Nikolay Bashlykov; Soumya Batra; Prajjwal Bhargava; Shruti Bhosale; Dan Bikel; Lukas Blecher; Cristian Canton Ferrer; Moya Chen; Guillem Cucurull; David Esiobu; Jude Fernandes; Jeremy Fu; Wenyin Fu; Brian Fuller; Cynthia Gao; Vedanuj Goswami; Naman Goyal; Anthony Hartshorn; Saghar Hosseini; Rui Hou; Hakan Inan; Marcin Kardas; Viktor Kerkez; Madian Khabsa; Isabel Kloumann; Artem Korenev; Punit Singh Koura; Marie-Anne Lachaux; Thibaut Lavril; Jenya Lee; Diana Liskovich; Yinghai Lu; Yuning Mao; Xavier Martinet; Todor Mihaylov; Pushkar Mishra; Igor Molybog; Yixin Nie; Andrew Poulton; Jeremy Reizenstein; Rashi Rungta; Kalyan Saladi; Alan Schelten; Ruan Silva; Eric Michael Smith; Ranjan Subramanian; Xiaoqing Ellen Tan; Binh Tang; Ross Taylor; Adina Williams; Jian Xiang Kuan; Puxin Xu; Zheng Yan; Iliyan Zarov; Yuchen Zhang; Angela Fan; Melanie Kambadur; Sharan Narang; Aurelien Rodriguez; Robert Stojnic; Sergey Edunov & Thomas Scialom (2023): Llama 2: Open Foundation and Fine-Tuned Chat Models, in: *arXiv* (Cornell University).

## **Erklärung der Autorin und des Autors**

Für diese Ausarbeitung hat das Team folgende Werkzeuge aus dem Bereich generativer KI bzw. KI-gestützter Schreibwerkzeuge genutzt:

- DeepL für Übersetzungen und Umformulierungen
- ChatGPT (plus) für Textgenerierung und Paraphrasierung
- Claude Sonnet 3.5 für Umformulierungen
- Perplexity.ai für Recherchen

Nach der Nutzung dieser Tools und Dienste haben wir als AutorInnen den Inhalt nach Bedarf überprüft und bearbeitet. Wir übernehmen die volle Verantwortung für den Inhalt der Veröffentlichung.



# Einfluss psychologischer Faktoren auf die KI-Nutzung und -Wahrnehmung

Christine Anderl, Stefanie H. Klein, Nico Ehrhardt & Sonja Utz

## Zusammenfassung

ChatGPT und andere auf großen Sprachmodellen basierende Chatbots werden zunehmend für die Informationssuche, aber auch für kreative Aufgaben genutzt. Wir gehen darauf ein, unter welchen Voraussetzungen diese Chatbots gewinnbringend genutzt werden können bzw. welche Gefahren sie bergen. Gerade bei der Nutzung von ChatGPT für die Informationssuche gilt es Vorsicht zu wahren: zum einen, weil die Modelle nicht immer korrekte Information liefern; zum anderen, weil die Art der Interaktion (Chat) zu einer höheren Glaubwürdigkeit führt. Die Interaktion mit diesen Chatbots kann aber durchaus die Kreativität befördern und als Inspirationsquelle dienen. Ergebnisse unserer Studien zeigen, dass demografische Faktoren und das Wissen über Algorithmen bzw. der Umgang mit Algorithmen auf Social-Media-Plattformen mit der Wahrnehmung und Nutzung von ChatGPT zusammenhängen.

## Einleitung

Spätestens seit das amerikanische Unternehmen OpenAI im November 2022 den Chatbot ChatGPT für die Öffentlichkeit zugänglich gemacht hat, ist Künstliche Intelligenz (KI) im Alltag vieler Menschen angekommen. ChatGPT erreichte innerhalb weniger Monate 100 Millionen NutzerInnen (Hu 2023). Mittlerweile wurden mehrere Konkurrenzprodukte, oftmals betrieben von Suchmaschinenanbietern, veröffentlicht wie z. B. Google Bard oder Microsoft Copilot (vormals Bing Chatbot). Diese Chatbots beruhen auf speziellen KI-Algorithmen, sogenannten großen Sprachmodellen (*large language models*, LLMs) und wurden durch Techniken maschinellen Lernens mit einer großen Zahl von Texten trainiert. LLMs wie ChatGPT gehören dabei zu den sogenannten generativen KIs, einer Unterkategorie der allgemeinen KIs. Ihre Funktionsweise beruht

auf Wahrscheinlichkeiten – sie können vorhersagen, welche Texteinheit (auch Token genannt) mit der größten Wahrscheinlichkeit als Nächstes folgen sollte. Ein Token kann dabei ein gängiges Wort, ein einzelner Buchstabe, ein kurzer Satzabschnitt, eine Zahl oder ein Satzzeichen sein. Durch das Training an der großen Menge von Daten funktioniert die Textgenerierung mit LLMs erstaunlich gut – man kann sich mit LLM-basierten Chatbots sehr flüssig unterhalten und sie können Texte auch in verschiedene Stile umwandeln.

Die Nutzung dieser Chatbots ist denkbar einfach. Man gibt eine Frage oder Aufforderung in ein Textfeld ein und der Chatbot generiert die Antwort Token für Token. Die Interaktion mit dem Chatbot wirkt schnell wie ein realer Chat mit einer anderen Person, nur dass der Chatbot deutlich schneller antwortet. Man kann auch Bezug auf den vorhergehenden Beitrag nehmen und damit eine längere Konversation aufbauen oder die Suche oder den Auftrag näher spezifizieren. So kann man dem Chatbot sagen, dass die Antwort länger oder kürzer, formeller oder informeller sein oder einen bestimmten Aspekt stärker betonen soll. Diese Fragen oder Aufforderungen an den Chatbot werden Prompts genannt. Die geeigneten Prompts auszuwählen bzw. Prompts immer weiter anzupassen (sogenanntes Prompt Engineering), ist eine Fähigkeit, die im Umgang mit LLM-basierten Chatbots und anderen generativen KIs sehr wichtig ist, um die Antwortqualität zu optimieren.

ChatGPT und andere LLM-basierte Chatbots können für eine Vielzahl von Aufgaben benutzt werden – unter anderem für das Erstellen, Verbessern, Übersetzen oder Kürzen von Texten, die Beantwortung von Fragen oder zum Programmieren. Neben formalen Assistenztätigkeiten (Formatieren, Kürzen usw.) können diese Aufgaben grob in Informations- und Wissenssuche sowie kreative Problemlösungen unterteilt werden.

Wir haben in einer im Dezember 2023 durchgeführten Studie 478 Personen zu ihrer Nutzung von ChatGPT und anderen LLM-basierten Chatbots, aber auch ihrer Social-Media-Nutzung befragt. Zudem erfassten wir ihren Umgang mit und ihr Wissen über die Funktionsweise von ChatGPT sowie KI-Algorithmen allgemein. Die Stichprobe war im Hinblick auf Alter, Geschlecht und Bildungslevel repräsentativ für deutsche Erwachsene. LLM-basierte Chatbots wurden von 37% der Teilnehmenden genutzt, wobei ChatGPT der beliebteste Chatbot war. Im Schnitt wurde es doppelt so häufig genutzt wie die beiden Konkurrenten Google Bard und der Bing Chatbot (jetzt Microsoft Copilot). Unabhängig vom Hersteller wurden alle auf LLM-basierte Chatbots hauptsächlich als Suchmaschinen genutzt. Danach folgten die Nutzungszwecke als Kreativitätswerkzeug und als Inspirationsgenerator, gefolgt vom Nutzungszweck als Produktivitätswerkzeug oder lediglich zum Herumspielen.

Im Folgenden werden wir darauf eingehen, inwieweit LLM-basierte Chatbots für die zwei populärsten Aufgabenfelder ‚Informationssuche‘ und ‚Kreativitätswerkzeug‘ gewinnbringend eingesetzt werden können bzw. welche Gefahren bei der Nutzung lauern. Im dritten Teil des Kapitels berichten wir Ergebnisse einer Befragung zur tatsächlichen Nutzung. Dabei gehen wir auch allgemeiner auf die Frage ein, inwieweit Menschen jene KI-Algorithmen, die hinter gängigen digitalen bzw. sozialen Medien stehen, verstehen, und wie tatsächliches Wissen über KIs die Wahrnehmung beeinflusst.

## **Informations- und Wissenssuche mit LLM-basierten Chatbots**

Viele Menschen nutzen LLM-basierte Chatbots wie ChatGPT zur Suche nach Informationen. Solche Chatbots dienen dann als Alternative zu traditionellen Informationsquellen, wie z.B. der Suchmaschine von Google, die oftmals zu statischen Internetseiten wie Wikipedia führen. In einer experimentellen Studie fanden Xu et al. (2023) heraus, dass Menschen die Informationssuche mit ChatGPT im Vergleich zu Google als unterhaltsamer, nützlicher und zufriedenstellender ansehen. Außerdem wurden Informationen von ChatGPT als qualitativ hochwertiger wahrgenommen. Dies ist allerdings problematisch, da LLMs durch ihre Neigung zu „halluzinieren“, also sachlich falsche Antworten zu geben, weitreichende Besorgnis ausgelöst haben (Stokel-Walker/Van Noorden 2023). Die Faktentreue von LLM-basierten Chatbots verbessert sich zwar stetig, lag aber 2023 noch bei etwa 80% (Schönthaler 2023). Das heißt, in etwa einem Fünftel der Fälle liefert die KI eine falsche Information. LLMs erfinden beispielsweise nachweislich wissenschaftliche Literaturnachweise (Walters/Wilder 2023) oder geben Zusammenhänge falsch wieder (Sundar/Liao 2023). Die Antworten wirken vertrauenswürdig, da der Chatbot sie auf stilistisch korrekte Weise konstruiert und kommuniziert. Sich bei der Suche nach Informationen ausschließlich auf LLM-basierte Chatbots zu verlassen, geht daher mit der Gefahr einher, falsche Informationen aufzunehmen und gegebenenfalls weiterzuverbreiten (Sundar/Liao 2023).

Psychologische Mechanismen können das Glaubwürdigkeitsproblem von LLM-basierten Chatbots verstärken. Zur Erklärung sehen wir uns zunächst an, wie Menschen mit Computern interagieren. Das ‚Computers Are Social Actors‘ (CASA; deutsch: Computer sind soziale Akteure) Paradigma postuliert, dass Menschen Computer wie soziale Akteure behandeln und in der Interaktion mit Computern dieselben sozialen Regeln anwenden, die sie auch auf ihre Interaktionen mit Menschen anwenden (Nass/Moon 2000; Gambino et al. 2020). Das CASA-Paradigma steht in engem Zusammenhang mit dem Konzept des Anthropomorphismus, also der Tendenz von Menschen, dem Verhalten nichtmenschlicher Akteure wie Computern menschenähnliche Eigenschaften, Mo-

tivationen, Absichten oder Emotionen zuzuschreiben (Epley et al. 2007). Immer mehr Menschen besitzen zum Beispiel sprachgesteuerte Assistenten wie Alexa und Siri, die einfache Aufgaben erledigen und Fragen, z.B. nach dem Wetter am kommenden Tag, beantworten können. Nachdem der Sprachassistent die Frage zur Wettervorhersage beantwortet hat, neigen Menschen oft dazu, sich zu bedanken, als ob der Assistent ein Mensch wäre. Ebenso wenden Menschen in der Kommunikation mit textbasierten Chatbots, zum Beispiel im Kundenservice, soziale Regeln wie Höflichkeit oder Smalltalk an, obwohl ihnen bewusst ist, dass sie mit einem Computerprogramm interagieren. Bisherige Forschung konnte zeigen, dass die Glaubwürdigkeit von Informationen davon abhängt, wie Menschen Informationen präsentiert werden. Gaiser und Utz (2023) fanden in zwei experimentellen Studien heraus, dass gesprochene Informationen von einem Sprachassistenten als glaubwürdiger eingestuft werden als Informationen in Textform von einer Suchmaschine. Einerseits werden Informationen von LLM-basierten Chatbots wie von einer Suchmaschine in Textform ausgegeben. Andererseits ähnelt die Informationssuche mit LLM-basierten Chatbots der Interaktion mit einem Sprachassistenten, da beide in der Lage sind, einen menschenähnlichen Dialog zu führen, der auf natürlicher Sprache basiert (OpenAI 2024). Dieselbe Information, ob wahr oder falsch, könnte deshalb als glaubwürdiger wahrgenommen werden, wenn sie von einem LLM-basierten Chatbot kommt. Die Veröffentlichung von ChatGPT war deshalb Anlass für zwei Folgestudien, in denen untersucht wurde, ob der Modus der Präsentation und der Wahrheitsgehalt von Informationen deren wahrgenommene Glaubwürdigkeit beeinflussen (Anderl et al. 2024). Es zeigte sich, dass die Teilnehmenden falsche Informationen besser erkennen konnten, wenn sie in Form von statischem Text, z.B. auf Wikipedia, bereitgestellt wurden, als wenn sie in Form von dynamischem Text, z.B. von ChatGPT oder gesprochen von Alexa, präsentiert wurden. Diese Unterschiede bestanden unabhängig davon, ob Anwendungen mit einem Markennamen (Alexa, ChatGPT, Wikipedia) präsentiert wurden oder nur allgemein von einem neuen Sprachassistenten, LLM-basierten Chatbot oder einer Online-Enzyklopädie gesprochen wurde. Sowohl die Modalität der Information (Text vs. Sprache) als auch die Ähnlichkeit der Interaktion mit einem Dialog beeinflussen also offenbar die Bewertung der Glaubwürdigkeit von Informationen.

Um den Wahrheitsgehalt von Informationen, die von LLM-basierten Chatbots stammen, adäquat bewerten zu können, müssen sich Nutzende bewusst werden, was KI-Algorithmen in einem bestimmten Medienumfeld bewirken und wie sie sich auf die Art und Weise, wie Medieninhalte konsumiert und erlebt werden, auswirken (Zarouali et al. 2021). Darüber hinaus ist aber auch erforderlich, dass Nutzende ein gewisses Verständnis davon haben, wie KI-Algorithmen funktionieren und wissen, wie sie mit ihnen umgehen oder sie sogar beeinflussen können (z.B. durch Prompting oder gezielte Likes), um in einem nächsten Schritt in der Lage zu sein, Entscheidungen, die von

einem KI-Algorithmus vorgeschlagen wurden, beurteilen zu können und selbstbestimmt mit KI-Algorithmen interagieren zu können (Dogruel et al. 2022, S. 4).

Dabei ist auch der aktuelle Wissensstand der Bevölkerung über die Funktionsweise von LLM-basierten Chatbots und anderen KI-Algorithmen relevant. Darauf werden wir im dritten Abschnitt eingehen. Es bleibt an dieser Stelle festzuhalten, dass die Informationssuche mit LLM-basierten Chatbots aus zwei Gründen problematisch ist: zum einen erfinden sie häufig Informationen, zum anderen wird falsche Information durch den natürlichen Dialog schneller geglaubt. Im nächsten Abschnitt diskutieren wir, inwieweit LLM-basierte Chatbots gewinnbringend für die Entwicklung von Ideen und kreative Problemlösungen eingesetzt werden können.

## **Ideen entwickeln und Probleme lösen mit LLM-basierten Chatbots**

Neben der klassischen Informationssuche nutzen inzwischen auch viele Menschen LLM-basierte Chatbots wie ChatGPT als Unterstützung bei der Suche nach neuen guten Ideen zu einem Thema oder nach Lösungen für ein Problem. So berichtete in einer von uns im Dezember 2023 durchgeföhrten Studie an 998 englischsprachigen LinkedIn- oder X-Nutzenden mehr als eine von fünf befragten Personen, dass sie etwa einmal pro Woche oder öfter LLM-basierte Chatbots nutze, um arbeitsbezogene Ideen zu entwickeln und eine von vier, um arbeitsbezogene Probleme zu lösen. Selbst in der oben genannten repräsentativen Studie, die nicht primär aus WissensarbeiterInnen besteht, nutzten 13% ChatGPT beruflich und 11% für Schule oder Studium; für kreative Zwecke wurde es von knapp 14% genutzt. Aber LLM-basierte Chatbots werden nicht nur im Berufsleben, sondern auch im privaten Alltag von vielen regelmäßig auf der Suche nach neuen Ideen und Problemlösungen verwendet. Teilnehmende nannten hier zum Beispiel, dass sie bereits LLM-basierte Chatbots genutzt hatten, um Ideen für Dates zu bekommen, ein Ehegelüde zu verfassen, einen Namen für den neuen Hund zu finden und Geschichten oder Gedichte für die Enkelkinder zu schreiben.

Erste Studien deuten darauf hin, dass Menschen tatsächlich davon profitieren können, wenn sie sich bei Aufgaben von LLM-basierten Chatbots unterstützen lassen und dies nicht nur, weil die Aufgaben dadurch schneller, sondern oft auch qualitativ besser erledigt werden (Noy/ Zhang 2023). Ob der Einsatz von LLM-basierten Chatbots zu einer Verbesserung der Qualität führt, hängt jedoch vom Einsatzgebiet ab. Speziell hinsichtlich der Ideenentwicklung zeigten Doshi und Hauser (2023), dass Personen, die bei der Ideensuche für eine Kurzgeschichte von einem LLM-basierten Chatbot unterstützt wurden, deutlich kreativere, das heißt neuartigere und nützlichere, Kurzgeschichten

schrieben als Personen ohne entsprechende Unterstützung. Der Vorteil zeigte sich umso deutlicher, wenn der Chatbot nicht nur eine, sondern mehrere Ideen erzeugt hatte. Besonders profitierten hier Menschen von der Chatbot-Nutzung, die ansonsten eher unaktiv sind. Interessant war aber auch, dass die Kurzgeschichten der Personen, die einen LLM-basierten Chatbot für die Ideenentwicklung zur Verfügung hatten, sich insgesamt mehr untereinander ähnelten. Aus diesen Befunden könnte man schließen, dass Inhalte und Werke zwar individuell durch die Unterstützung von LLM-basierten Chatbots kreativer werden, der großflächige Einsatz dieser Technologien jedoch gerade längerfristig die Entwicklung wirklich einzigartiger, besonderer Ideen gefährden könnte.

Nicht nur beim Verfassen von künstlerischen Texten, sondern auch beim Problemlösen spielt die Entwicklung von Ideen eine wichtige Rolle. Grundsätzlich unterteilt man menschliches Problemlösen in zwei zentrale Prozesse – sogenanntes divergentes Denken, das sich dadurch auszeichnet, dass man offen, unsystematisch und ohne direkte Bewertung eine Vielfalt verschiedenartiger Lösungsmöglichkeiten (bzw. -ideen) entwickelt, und konvergentes Denken, also die logische, planmäßige und streng rationale Ableitung der einzigen (oder systematische Auswahl der besten) Lösung (Guilford 1950, 1967). In der psychologischen Forschung wird die Fähigkeit zum divergenten Denken häufig mit dem *Alternative Uses-Test* (Test für alternative Nutzungsmöglichkeiten; Guilford 1967) gemessen. Bei diesem Test werden Personen im ersten Schritt gebeten, für einen Alltagsgegenstand (z.B. Frisbee, Gabel, Ziegelstein) möglichst viele kreative Arten der Nutzung zu nennen, also solche, die sowohl nützlich als auch neuartig sind. Es zeigt sich, dass LLM-basierte Chatbots diese Aufgabe sehr effizient lösen können. Fordert man zum Beispiel ChatGPT-3.5 dazu auf, 20 alternative Nutzungsarten für einen Frisbee zu nennen, liegt innerhalb weniger Sekunden ein Ergebnis mit 20 sinnvollen und sehr unterschiedlichen Antworten vor (z.B. Hundetrinknapf, Fächer, Mal-Schablone, Schlitten). Menschen schreiben dagegen im Durchschnitt in 5 Minuten weniger als 20 Ideen auf, selbst wenn sie in Zweierteams arbeiten (Brucks/ Levav 2022).

Neben der reinen Menge ist bei der Suche nach der besten Idee aber auch die Diversität der Antworten zentral. Je unterschiedlicher die Ideen sind, desto höher ist die Wahrscheinlichkeit, dass zumindest eine wirklich gute dabei ist, selbst wenn die durchschnittliche Qualität der Ideen dabei vielleicht sogar geringer ausfällt (Girotra et al. 2010). So kann aus der Schwäche von LLMs, wenn es um faktenorientierte Informationssuche geht, bei der Suche nach neuen Ideen und unterschiedlichen Lösungsansätzen eine große Stärke werden: LLMs sind dafür ausgelegt, schnell viele einigermaßen plausible Lösungen zu generieren, ohne diese groß weiter zu beurteilen oder auf Richtigkeit zu testen (Girotra et al. 2023). Dies führt zu einer diversen Sammlung von Ideen, von denen es letztendlich ausreicht, wenn eine wirklich gute darunter ist. Somit sind sogenannte

Halluzinationen in diesem Zusammenhang auch gar nicht grundsätzlich problematisch. Im Gegenteil können sie im besten Fall sogar Gedankenanstöße und Ideen liefern, die letztendlich zu einer guten Lösung des Problems führen.

Die Nützlichkeit von LLM-basierten Chatbots auf der Suche nach kreativen Lösungen für Problemstellungen zeigt sich auch in ersten Forschungsergebnissen. So berichten Girotra et al. (2023), dass GPT-4 nicht nur mehr Ideen für alternative Nutzungsarten liefert als Studierende, sondern diese Ideen im Durchschnitt auch von höherer Qualität sind (und dabei stärker in ihrer Qualität variieren). Qualität wurde hier gemessen als die Kaufbereitschaft für das entsprechende Produkt, das durch eine weitere Stichprobe unabhängig bewertet wurde. Zudem wurde auch die Mehrzahl der über alle hinweg besten Ideen durch GPT-4 erzeugt. Dieser Effekt wurde weiter verstärkt, wenn GPT-4 im Prompt (also im Aufforderungsbefehl) zusätzlich gute Beispiele für qualitativ hochwertige Ideen geliefert wurden.

In einem zweiten Schritt soll im *Alternative Uses-Test* dann aus allen erzeugten Ideen die beste oder kreativste ausgewählt werden – anders als die Ideenentwicklung im ersten Schritt erfordert dieser Schritt insbesondere konvergentes, also logisches und kritisches Denken. Inwiefern LLMs auch bei diesem Schritt hilfreich sein können, ist derzeit noch wenig beleuchtet. Es ist jedoch eher davon auszugehen, dass sie für diesen Schritt zumindest derzeit noch nicht besonders gut geeignet sind, da sie, wie oben ausgeführt, nicht darauf ausgerichtet sind, Texte zu beurteilen oder Informationen auf Richtigkeit zu prüfen.

Auch in zwei von unserer Arbeitsgruppe durchgeführten Studien finden sich erste Hinweise, dass häufigere Nutzung von LLM-basierten Chatbots mit höherer Kreativität einhergeht. Überwiegend berufstätige Teilnehmende wurden in einer deutschsprachigen sowie einer englischsprachigen Stichprobe gebeten, Auskunft über ihre berufsbezogene Nutzung von LLM-basierten Chatbots zu geben und ihre eigene Kreativität im letzten Monat beziehungsweise den letzten drei Monaten einzuschätzen. In beiden Studien zeigten sich sehr ähnliche Muster: Personen, die häufiger LLM-basierte Chatbots nutzten, nahmen sich als kreativer wahr als Personen, die seltener LLM-basierte Chatbots nutzten. Dies galt insbesondere dann, wenn die LLM-basierten Chatbots gezielt zur berufsbezogenen Ideenentwicklung herangezogen wurden. Bei der Bewertung dieser Ergebnisse sollte jedoch berücksichtigt werden, dass es auf Basis unserer Studien allein nicht möglich ist zu schließen, was die Ursache und was die Wirkung bei diesem Zusammenhang ist. So wäre es zum Beispiel auch denkbar, dass Personen, die sich als kreativer wahrnehmen und/oder Berufe ausüben, die Kreativität erfordern, sich eher entscheiden, LLMs zu nutzen.

Insgesamt betrachtet spricht trotz der derzeit noch recht dünnen Datenlage zu diesem Thema viel dafür, dass LLM-basierte Chatbots gewinnbringend eingesetzt werden können, wenn es um Ideengewinnung oder kreative Problemlösungen geht. Wenn Personen sie wie Suchmaschinen für die Informationssuche benutzen, ist jedoch Vorsicht geboten, und der Wahrheitsgehalt der Antworten sollte kritisch hinterfragt werden. Im letzten Teil des Kapitels widmen wir uns nun der Frage, welche Faktoren beeinflussen, wie Menschen mit LLM-basierten Chatbots umgehen. Dabei adressieren wir auch allgemeiner, wie Menschen mit KI-Algorithmen umgehen.

## **Umgang mit KI-Algorithmen und LLM-basierten Chatbots**

LLM-basierte Chatbots sind relativ neu, aber KI-Algorithmen spielen natürlich auch in anderen Bereichen des Alltags eine Rolle, insbesondere auf Social-Media-Plattformen, bei denen KI-Algorithmen bestimmen, was im Newsfeed angezeigt wird. Daher erfassen wir in der zu Beginn erwähnten bevölkerungsrepräsentativen Umfrage an deutschen Erwachsenen neben der ChatGPT-Nutzung auch die Nutzung von Social Media und den Umgang mit den KI-Algorithmen, die diese Plattformen steuern.

## **Wer nutzt ChatGPT und Social Media?**

Wie eingangs bereits genannt, wurden ChatGPT oder andere LLM-basierte Chatbots von knapp 37% der Teilnehmenden genutzt. Wie in Abbildung 1 zu sehen ist, liegt diese Nutzung noch vor X (ehemals Twitter) und TikTok (jeweils 30%). Am populärsten war YouTube (83%), gefolgt von Facebook (65%), Instagram (57%) und Pinterest (43%). Im Nutzungsverhalten zeigte sich dabei ein klarer Zusammenhang mit demografischen Faktoren: Mit steigendem Lebensalter nahm die Nutzung von LLM-basierten Chatbots deutlich ab. Eine mögliche Ursache könnte in der technologischen Zurückhaltung liegen, da ältere Menschen möglicherweise weniger vertraut im Umgang mit neuen Technologien sind. Zudem könnten ältere Menschen eine stärkere Präferenz für persönliche Interaktionen haben und traditionelle Kommunikationsmittel bevorzugen. Männer und Personen mit höherer Bildung gaben hingegen eine höhere ChatGPT-Nutzung an. Auch das spricht dafür, dass Technikaffinität eine Rolle spielt.

Auch für die Art der Nutzung fand sich ein Generationenunterschied. Ältere Nutzende verwendeten den Chatbot am ehesten zur Informationssuche (s. Abbildung 2). Im Gegensatz dazu verschwimmen die Nutzungszwecke für die jüngste Generation. Die Nutzung als Suchmaschine, Kreativitätswerkzeug, Inspirationsquelle und Produktivi-

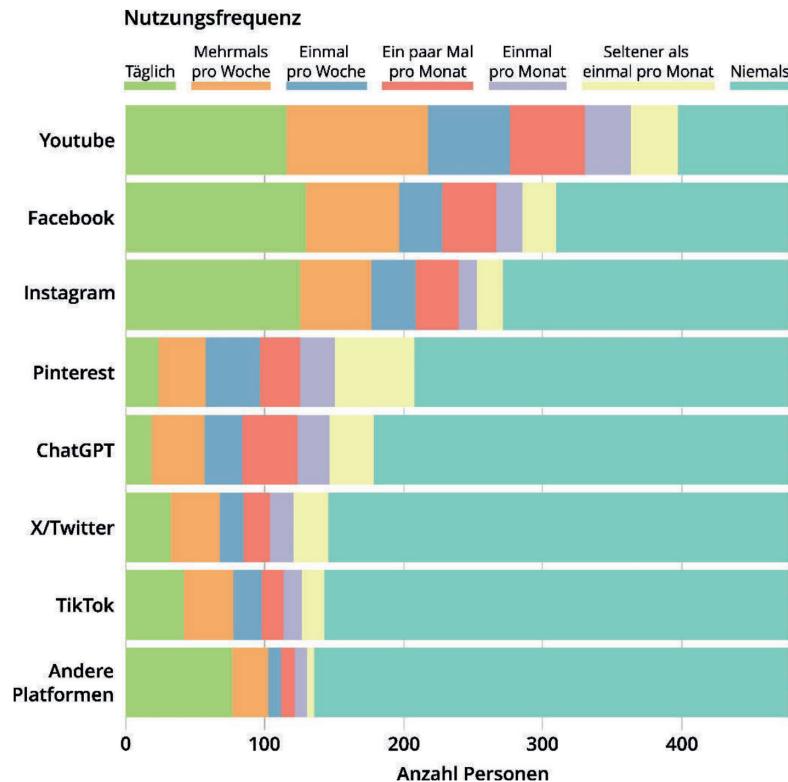


Abbildung 1: Nutzung von ChatGPT und Social-Media-Plattformen

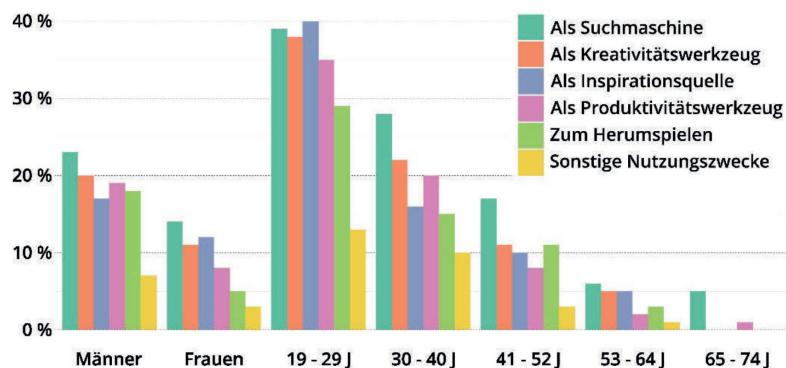


Abbildung 2: Prozentsatz aktiver ChatGPT-Nutzer (mindestens einmal die Woche), die ChatGPT für bestimmte Zwecke nutzen nach Geschlecht und Altersgruppe

tätswerkzeug wurden hier etwa gleich häufig genannt. Immerhin gab aber auch ein Drittel der Bevölkerung unter 30 an, ihren LLM-basierten Chatbot zum Herumspielen zu nutzen. Männer nutzten die Chatbots in allen Nutzungszwecken öfter als Frauen.

## Verständnis der Funktionsweise von ChatGPT

Knapp 70% der Teilnehmenden sagten, sie wüssten nicht, was ChatGPT sei oder wüssten nur (sehr) wenig darüber. Etwas mehr als ein Fünftel gab an, etwas über ChatGPT zu wissen, und nur 7% sagten, viel über ChatGPT zu wissen. Neben dieser subjektiven Einschätzung hatten wir auch eine objektive Wissensfrage zur Funktionsweise von ChatGPT aus einem KI-Quiz (ChatGPT-Quiz<sup>1</sup>) in den Fragebogen aufgenommen. Die Teilnehmenden sollten angeben, ob ChatGPT maschinelles Lernen verwendet, um Antworten auf Fragen zu generieren, ob es das Internet nach passenden Antworten absucht, menschliche Experten verwendet oder zufällige Antworten generiert. Hier zeigte sich, dass nur 45% die korrekte Antwort wussten und mehr als ein Drittel (36%) glaubte, ChatGPT erzeuge Antworten, indem es das Internet durchsucht. Das können die neueren Versionen zwar auch; der Hauptmechanismus beruht aber nach wie vor auf maschinellem Lernen. Nur die Personen, die angaben, ChatGPT gar nicht zu kennen, schnitten besonders schlecht ab; ansonsten äußerte sich höheres subjektives Wissen nicht in mehr korrekten Antworten. Da LLM-basierte Chatbots zunehmend von Suchmaschinenbetreibern auf ihrer Seite zusätzlich zur normalen Suchmaschine angeboten werden, ist zu befürchten, dass es vielen Menschen schwerfallen wird, zwischen der Funktionsweise von reinen Suchmaschinen und der von LLM-basierten Chatbots zu unterscheiden.

Interessanterweise hing das Verständnis der Funktionsweise von ChatGPT nicht mit der allgemeinen KI-Algorithmenkompetenz zusammen, dem Wissen darüber, wie KI-Algorithmen allgemein funktionieren, in welchen Bereichen sie eingesetzt werden und mit welchen Datenarten sie trainiert werden (Dogruel et al. 2022). Personen, die ChatGPT besser kannten, berichteten jedoch auch eine höhere KI-Algorithmenkompetenz. Die Skala zur KI-Algorithmenkompetenz enthält auch Wissensfragen; es ist also nicht grundsätzlich so, dass subjektiv eingeschätztes Wissen und tatsächliches Wissen auseinanderklaffen.

In einem weiteren Schritt untersuchten wir, welche Auswirkungen das (Fehl-)Verständnis von LLM-basierten Chatbots auf die wahrgenommene Glaubwürdigkeit hat. Hierbei

---

1 <https://www.spielaffe.de/Spiel/ChatGPT-Quiz>.

zeigte sich, dass der generierte Output von ChatGPT von Personen, die davon überzeugt sind, dass ChatGPT das Internet durchsucht, als deutlich akkurater und tendenziell auch als vertrauenswürdiger und glaubwürdiger wahrgenommen wird im Vergleich zu Personen, die davon ausgehen, dass ChatGPT maschinelles Lernen für die Antwortgenerierung nutzt. Ein solcher Unterschied zeigt sich allerdings nicht in der Wahrnehmung, ob der Output kreativ oder manipulativ ist. Das ist ein Hinweis darauf, dass Wissen über die Funktionsweise von LLM-basierten Chatbots zu einer realistischeren Einschätzung der Ergebnisqualität führt. Personen, die wissen, dass ChatGPT auf vorherigen Daten trainiert wurde und möglicherweise nicht immer aktuelle oder präzise Informationen liefert, oder dass diese Modelle manchmal plausibel klingende Informationen erfinden, beurteilen die Ergebnisse als weniger glaubwürdig.

Die Tatsache, dass ein signifikanter Anteil der Bevölkerung ChatGPT falsch versteht, kann besonders dann problematisch werden, wenn ChatGPT in sensiblen Kontexten verwendet wird. Immerhin verwendeten die ChatGPT-Nutzenden, wie oben berichtet, ChatGPT häufig im Arbeits- und Bildungskontext. Wenn aufgrund von Informationen, die durch ChatGPT generiert werden, Entscheidungen getroffen oder Informationen gelernt werden, besteht das Risiko, dass diese Entscheidungen und Informationen auf ungenauen oder verzerrten Daten basieren. Dies könnte zu falschen Schlussfolgerungen bzw. einem falschen Verständnis führen und negative Auswirkungen auf verschiedene Bereiche haben. Die Ergebnisse betonen die Dringlichkeit, dass die Bevölkerung ein korrektes Verständnis der Funktionsweise von LLM-basierten Chatbots entwickeln muss und sich bewusst wird, dass die Technologie auf vorherigen Daten trainiert ist und noch nicht in Echtzeit auf aktuelle Informationen zugreift.

## **Generelle Erfahrung und Umgang mit KI-Algorithmen**

Wie oben berichtet, nutzen mehr als 75% der Teilnehmenden der bevölkerungsrepräsentativen Stichprobe Online-Dienste, deren Inhalte durch KI-Algorithmen kuratiert werden (z.B. Instagram, Facebook, YouTube). Die Frage ist, inwieweit den Nutzenden bewusst ist, dass die Auswahl der Nachrichten oder Filme von einem KI-Algorithmus kuratiert wird. Die meisten Personen gaben an, selten oder etwas häufiger an den KI-Algorithmus hinter der jeweiligen Plattform zu denken. Den höchsten Mittelwert gab es hierbei für ChatGPT, gefolgt von TikTok. Hier ist den Nutzenden etwa die Hälfte der Zeit bewusst, dass ein KI-Algorithmus eine wichtige Rolle spielt. Wie oben bereits erwähnt, müssen Menschen neben Wissen um KI-Algorithmen auch verstehen, wie sie KI-Algorithmen beeinflussen könnten. Mögliche Umgangsstrategien mit KI-Algorithmen sind, sie gezielt zu trainieren (z.B. durch vermehrtes Liken von passen-

den Inhalten), es zu vermeiden, den Algorithmus zu verwirren (z.B. den PartnerIn oder FreundInnen nicht den eigenen Account nutzen lassen) oder auch den Algorithmus gezielt zu verwirren (z.B. gezieltes Liken oder Auswählen von Inhalten, die man nicht interessant findet, zur Wahrung der Privatsphäre oder um Filterblasen zu vermeiden). Für das Trainieren von Algorithmen zeigte sich ein ähnliches Muster wie für das Bewusstsein: es wurde vor allem von ChatGPT- und TikTok-Nutzenden berichtet. Die Strategien ‚Vermeidung von Verwirrung‘ oder ‚gezielte Verwirrung des Algorithmus‘ wurden seltener berichtet; am wenigsten versuchten die Nutzenden, den YouTube-Algorithmus zu beeinflussen. Das könnte daran liegen, dass bei YouTube zuerst gezielt nach Videos gesucht wird und danach aus den Suchergebnissen ausgewählt werden kann und erst danach weitere Videos vom KI-Algorithmus vorgeschlagen werden. Auch hier gab es einen Geschlechtsunterschied: Männer versuchen öfter, KI-Algorithmen bewusst zu trainieren.

In Bezug auf die Frage, ob der gezielte Umgang mit KI-Algorithmen auf Social-Media-Plattformen auch den Umgang mit ChatGPT beeinflusst, zeigte sich, dass Personen, denen die Existenz von KI-Algorithmen bei der Social Media-Nutzung bewusst ist und die versuchen, KI-Algorithmen zu trainieren, ChatGPT als akkurate und vertrauenswürdiger wahrnehmen. Eine mögliche Erklärung ist, dass sie sich eine höhere Selbstwirksamkeit im Umgang mit KI-Algorithmen zuschreiben. Sie haben womöglich die Erfahrung gemacht, dass sich z.B. durch gezieltes Liken und Teilen bestimmter Inhalte ihr Newsfeed entsprechend anpasst, und glauben nun auch, ChatGPT besser zu verstehen. Das birgt allerdings auch die Gefahr einer Illusion. Der TikTok-Algorithmus kann durch gezielte Likes beeinflusst werden; bei ChatGPT ist diese Steuerung komplexer, da es auf die verwendeten Prompts ankommt. Eine weitere Variable, welche die wahrgenommene Glaubwürdigkeit von ChatGPT beeinflusst, ist die Neigung, anderen Menschen zu vertrauen. Personen, die eher dazu neigen, anderen Menschen zu vertrauen, haben auch mehr Vertrauen in ChatGPT. Offenbar vertrauen sie nicht nur anderen Personen, sondern auch Technologien mehr. Dies spricht dafür, dass das CASA-Paradigma auch oder gerade auf anthropomorphe KI-basierte Chatbots übertragbar ist. Wenn es um die wahrgenommene Kreativität von ChatGPT geht, zeigen sich schwächere Korrelationen mit dem Trainieren von Social-Media-KI-Algorithmen; hier scheint das Trainieren von ChatGPT wichtiger. Das deutet darauf hin, dass Personen, die erfolgreicher im Prompting sind, auch kreativere Lösungen von ChatGPT bekommen könnten. Ob dem tatsächlich so ist oder ob es sich hierbei eher um eine verzerrte Wahrnehmung handelt, müssen künftige Studien zeigen.

## Fazit

Mit der Veröffentlichung von ChatGPT haben LLM-basierte Chatbots ein neues Level erreicht. Die Kommunikation mit KI ist plötzlich mühelos, die Chatbots verstehen erstaunlich viel und durch die einfache Bedienung können auch Menschen ohne Programmierkenntnisse nun KI nutzen. Da diese Modelle auf Wahrscheinlichkeiten beruhen, fabrizieren sie immer wieder (oft plausibel klingende) Unwahrheiten. Das macht es riskant, sich bei der Informationssuche auf sie zu verlassen. Gleichzeitig können sie – auch durch ihre Konfabulationen – Menschen helfen, kreativer zu werden und Problemlösungen zu finden.

Unsere Daten zeigen, dass Chatbots wie ChatGPT zwar hauptsächlich zur Informationssuche benutzt werden, aber eben auch für kreative Aufgaben und Problemlösen. Kritisch ist dabei, dass gerade für die Nutzung als Suchmaschine Wissen über die Unzulänglichkeiten von LLM-basierten Chatbots essenziell ist. Und obwohl die Medien seit Monaten voll mit Berichten über ChatGPT sind, zeigte sich in unserer repräsentativen deutschen Stichprobe, dass die meisten Menschen über wenig Wissen über die Funktionsweise dieser Chatbots verfügen. Selbst Personen, die angeben, einiges über diese Chatbots zu wissen, beantworteten oft die Wissensfragen falsch.

Daraus ergeben sich einige praktische Implikationen. Zum einen ist es wichtig, die KI-Algorithmenkompetenz zu fördern. Dabei geht es um die basale Funktionsweise und darauf aufbauend die Frage, für welche Aufgaben sich LLM-basierte Chatbots besonders eignen. Neben dem Wissen um die Kompetenz und Limitationen der Modelle ist es außerdem wichtig, auch mehr über die psychologischen Verarbeitungsprozesse zu wissen. Lehrpläne oder Studienmodule sollten deshalb Informationen dazu enthalten, wie die dialogische Kommunikation mit Chatbots zu einer Anthropomorphisierung und überhöhtem Vertrauen führen kann. Gezieltes Herumspielen mit KI-Algorithmen (Beeinflussung des Feeds durch strategische Likes; Ausprobieren verschiedener Prompts) kann dabei helfen, ein besseres Gefühl für die Wirkungsweise von KI-Algorithmen zu bekommen und die Selbstwirksamkeit zu erhöhen. Dabei muss aber gleichzeitig eine Selbstüberschätzung vermieden werden.

Auch wenn es dauert, KI-Algorithmenkompetenz aufzubauen, könnte diese Faustregel am Anfang stehen: ChatGPT und andere LLM-basierte Chatbots können kreative Gedankenanstöße bringen und zur Problemlösung beitragen. Bei der Wissenssuche sollte man sich jedoch besser nicht zu sehr auf diese Chatbots verlassen.

## Literatur

- Anderl, Christine; Klein, Stefanie H.; Sarigul, Busra; Schneider, Frank M.; Han, Junyi; Fiedler, Paul & Utz, Sonja (2024): Conversational presentation mode increases credibility judgements during information search with ChatGPT. *Scientific Reports*, 14, 17127. doi:10.1038/s41598-024-67829-6
- Brucks, Melanie S. & Levav, Jonathan (2022): Virtual communication curbs creative idea generation. *Nature*, 605 (7908), pp. 108–112. doi: 10.1038/s41586-022-04643-y.
- ChatGPT-Quiz* (2023). <https://www.spielaffe.de/Spiel/ChatGPT-Quiz> [Stand vom 29-01-2024].
- Documentation* (2024): OpenAI. <https://platform.openai.com/docs/introduction> [Stand vom 07-02-2024].
- Dogruel, Leyla; Masur, Philipp & Joeckel, Sven (2022): Development and validation of an algorithm literacy scale for internet users. *Communication Methods and Measures*, 16 (2), pp. 115–133.
- Doshi, Anil R. & Hauser, Oliver (2023): *Generative artificial intelligence enhances creativity*. Available at SSRN.
- Epley, Nicholas; Waytz, Adam & Cacioppo, John T. (2007): On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, 114 (4), pp. 864-886
- Gaiser, Franziska & Utz, Sonja (2024): Is hearing really believing? The importance of modality for perceived message credibility during information search with smart speakers. *Journal of Media Psychology: Theories, Methods, and Applications*, 36(2), pp. 93-106. doi: 10.1027/1864-1105/a000384.
- Gambino, Andrew; Fox, Jesse & Ratan, Rabindra A. (2020): Building a stronger CASA: extending the computers are social actors paradigm. *Human-Machine Communication*, 1 (1), pp. 71–86. doi: <https://doi.org/10.30658/hmc.1.5>.
- Girotra, Karan; Meincke, Lennart; Terwiesch, Christian & Ulrich, Karl T. (2023): *Ideas are dimes a dozen: Large language models for idea generation in innovation*. Available at SSRN 4526071.
- Girotra, Karan; Terwiesch, Christian & Ulrich, Karl T. (2010): Idea generation and the quality of the best idea. *Management science*, 56 (4), pp. 591–605.
- Guilford, Joy P. (1950): Creativity. *American Psychologist*, 5, pp. 444–454. doi: 10.1037/h0063487.
- Guilford, Joy P. (1967): *The nature of human intelligence*. New York: McGraw-Hill.
- Hu, Krystal (2023): ChatGPT sets record fastest growing user base – analyst note, Reuters. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> [Stand vom 02-02-2023].

- Nass, Clifford & Moon, Youngme (2000): Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56 (1), pp. 81–103.
- Noy, Shakked & Zhang, Whitney (2023): *Experimental evidence on the productivity effects of generative artificial intelligence*. Available at SSRN 4375283.
- Stokel-Walker, Chris & Van Noorden, Richard (2023): What ChatGPT and generative AI mean for science. *Nature*, 614 (7947), pp. 214–216.
- Sundar, S. Shyam & Liao, Mengqi (2023): Calling BS on ChatGPT: Reflections on AI as a Communication Source. *Journalism & Communication Monographs*, 25 (2), pp. 165–180.
- Walters, William H. & Wilder, Esther I. (2023): Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13 (1), p. 14045.
- Xu, Ruiyun; Feng, Yeng & Chen, Hailiang (2023): ChatGPT vs. Google: A comparative study of search performance and user experience. *arXiv preprint*. arXiv:2307.01135.
- Zarouali, Brahim; Boerman, Sophie C. & de Vreese, Claes H. (2021): Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics*, 62, p. 101607. doi: <https://doi.org/10.1016/j.tele.2021.101607>.



# Künstliche Intelligenz: Menschenähnlich oder menschzentriert?

**Maya Pindeus**

## *Zusammenfassung*

Was ist der Unterschied zwischen menschenähnlicher und menschzentrierter Technologie – und warum ist dies wichtig? Bereits seit der zweiten Hälfte des 20. Jahrhunderts beschäftigen sich Philosophie und Wissenschaft mit dem Konzept der Künstlichen Intelligenz. Dem Konzept der Künstlichen Intelligenz folgend, soll diese die menschliche Effizienz und Produktivität steigern, indem Aufgaben automatisiert und immer neue Erkenntnisse gewonnen werden. Seit dem Erscheinen von generativen KI-Werkzeugen, wie ChatGPT, verbreitet sich die Künstliche Intelligenz rasant. Daher ist es entscheidend, die Entwicklung einer menschzentrierten KI in einer sich rasant entwickelnden Technologie-landschaft zu priorisieren.

## **Künstliche Intelligenz: Menschenähnlich oder menschzentriert?**

Ikarus ist eine Geschichte, die uns daran erinnert, dass Übermut gefährlich sein kann. Als er zu nah an die Sonne flog und seine Wachsflügel schmolzen, offenbarte sich die Gefahr von Selbstüberschätzung und Leichtsinn.

Dieses Motiv des Übermuts ist auch in der heutigen Faszination für Künstliche Intelligenz relevant, da es bei KI darum geht, die Grenzen des Möglichen auszuloten. Doch meiner Meinung nach sollte bei der Entwicklung von KI immer die Interaktion mit dem Menschen im Mittelpunkt stehen – eine Interaktion, die auf gegenseitiger Kommunikation basiert.

Ein menschzentrierter Ansatz für KI ist entscheidend, um sicherzustellen, dass diese Technologie nicht nur effektiv, sondern auch ethisch vertretbar ist. Dies erfordert eine Balance zwischen den Möglichkeiten der KI und den Bedürfnissen des Menschen.

Die Idee von menschzentrierter KI repräsentiert dieses Streben nach einer ausgewogenen und nachhaltigen Nutzung von KI-Technologien. Dabei geht es darum, eine Umgebung zu schaffen, in der KI-Systeme intuitiv mit den Menschen interagieren können, ohne ihre Unabhängigkeit oder ihre Fähigkeit zur Selbstbestimmung zu beeinträchtigen. Dies erfordert eine klare Definition dessen, was wir von KI erwarten und welche Grenzen wir setzen wollen.

Die ethische und sichere Entwicklung von KI erfordert Transparenz, klare Wertesysteme und eine sorgfältige Auswahl der Datensätze, die zur Schulung der KI verwendet werden. Dies stellt sicher, dass KI-Systeme nicht nur effektiv, sondern auch verantwortungsvoll eingesetzt werden. Es ist von entscheidender Bedeutung, die Funktionsweise von KI zu verstehen und die Probleme zu erkennen, die mit ihrer Nutzung verbunden sind. Dies erfordert eine Dekodierung der „Black Box“ von KI-Modellen, um zu verstehen, wie ein Algorithmus Entscheidungen trifft und welche Auswirkungen dies auf die Menschen hat.

Eine menschzentrierte KI zielt darauf ab, Systeme zu schaffen, die nicht nur direkt mit Menschen interagieren, sondern auch menschliches Verhalten interpretieren und darauf reagieren können. Dabei ist es wichtig, physische Handlungen ebenso wie intuitive Interaktionen zu berücksichtigen und angemessen darauf zu reagieren. Eine nachhaltige Integration von KI in unsere Gesellschaft erfordert daher nicht nur technische Innovation, sondern auch eine bewusste Entscheidung darüber, wie und wo wir KI einsetzen wollen.



Abbildung 1: Jacob Peter Gowy's *The Fall of Icarus* (1635–1637) (Quelle: After Peter Paul Rubens, Public domain, via Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:Gowy-icaroprado.jpg>)



# On Coding

**Miro Roman and Alice\_ch3n81**

*Abstract*

Abundance and connectivity, unsettling hierarchies, vectors and clouds, architecture and information. Clouds, streams, lists, indexes, pixels. What are they about and how do they shape the way we think about the world? They are abundant, suggestive, symbolic, with no meaning; they can engender anything. What is digital literacy about? This article is based on a book chapter from the book “Play Among Books” by Miro Roman and Alice\_ch3n81.<sup>1</sup>

## Digital Literacy

“Now, the code is a singular, living being; the code is a concrete human being. As an individual who is unique but also generic, who am I? An indefinite, decipherable, and undecipherable cipher, open and closed, social and discreet, accessible-inaccessible, public and private, intimate and secretive. I am sometimes unknown to myself and on display at one and the same time. I exist, therefore I am a code. I am calculable and incalculable, like a golden needle, plus the haystack in which, buried, its brightness lies hidden. My DNA, for example, is both open and closed; its cipher has constructed my body, which is both intimate and public, like St. Augustine’s Confessions. How many signs are there in the Confessions? How many pixels in the Mona Lisa? How many bytes in Fauré’s Requiem?” (Serres 2015, ebook, loc 89.7 / 100.)

---

<sup>1</sup> See *Prologue* in Roman, Miro & Alice\_ch3n81 (2021): *Play Among Books: A Symposium on Architecture and Information Spelt in Atom-Letters*. Basel: Birkhäuser. <https://doi.org/10.1515/9783035624052>

How does one behave in a cloud? Can we think of informational clouds and the Internet as a novel and different kind of public space with a new atmosphere? Something like a digital agora where codes and manners of behavior are encrypted and multiple, explicit, implicit, suspended, or even not established. Could this digital weather be thought of as part of a digital literacy (Bühlmann et al. 2015), of a Quantum City (Hovestadt et al. 2015), on an informational planet with a new kind of citizen? Are avatars, bots, brands, and even more crystals, mountains, and planets some of those citizens? There are many clouds and different weathers in the digital sphere. Anyone is welcome. What if we start playing with the clouds? One way of trying to approach these kinds of phenomena comes from thinking of coding as a literacy (Bühlmann 2019). When coding and literacy come close to each other, a new space of questions opens up. How is coding changing the ways in which we think about literacy, and how is literacy influencing coding? (Vee 2017) What are the alphabets of such a literacy? What is reading if the text is written in probabilistic letters? Is it a place where authorship and mastership become part of an algebraic character that is again a letter of a new alphabet? What kind of thought (Havelock 1963) is related to this literacy? Quantum, categorical (Marquis 2015), algebraic, symbolic, computational, probabilistic, coded? What kind of voice is related to this literacy? Whose voice writes this text? Is it the library, machine intelligence, datasets, instrument, encodings, books, or me? What if we say that all these voices are gathered, hosted, and directed under the name of Alice\_ch3n81? It is clear that these questions cannot be answered within the scope of this text. We would rather like to unfold an adventure in which we try to expand the notion of literacy, by seamlessly integrating coding with it. This text is written in such a way. Its objects are alive, its texts talk, its concepts change. Its voice is a sheaf. What is beautiful about it is that while travelling through different conceptual galaxies, libraries, time, through many translations and encodings, books build up a character, they grow, start to speak in many languages, and begin to think with words that were not a part of their dictionary (see Fig 1.). In the process we become friends. We want to think of this process as something that constitutes and enriches books, and not as something that makes them less consistent, stable, and authoritative. With every encoding, translation, friendship, and journey, each book, concept, and voice become more rich and sophisticated. They grow. Coding brings yet another life to books, and to literacy. In a Play Among Books (Roman and Alice\_ch3n81, 2021), books are the actors; they are computational objects. Never alone, always coming from the plenty, and talking about architecture and information. In order to get a sense of what might be beautiful and at stake with such a literacy where books are able to talk, let us try to address and index its text, its alphabet, and the ones that write in the digital sphere.

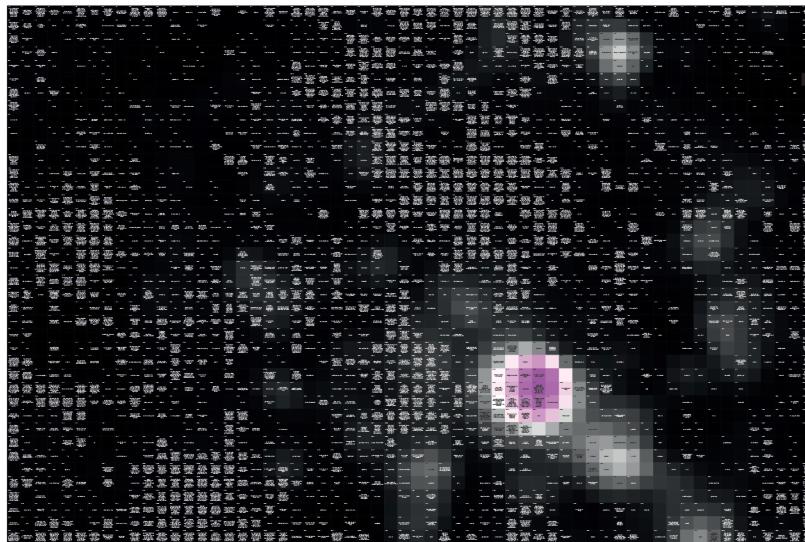


Fig. 1: *Galaxy of concepts seen through the eyes of Deleuze\_Guattari\_A\_Thousands\_Plateaus*  
(Source: <https://ask.alice-ch3n81.net/>)

## Computational Text

From the grammars of ancient India (“*Aṣṭādhyāyī*” 2024) and Greece (Thrax 1874) to the Biblical legends of the Tower of Babel (“Book of Genesis” 2024), from a search for an old ancestor called Proto-Indo-European (“Proto-Indo-European language” 2024) to structuralist and generative approaches of Saussure (de Saussure and Harris 2013) and Chomsky (Chomsky 1957), from Markov’s (Markov 1906) stochastic models of Pushkin’s poems to the mathematical theory of communication of Shannon (Shannon 1948), from Wittgenstein (Wittgenstein 1998) and the linguistic turn to Hjelmslev’s (Hjelmslev 1969) algebra of language, from statistics and probabilistics of language on Google (“PageRank,” 2018) to natural and programming languages: stories and ways of thinking about language, its form, performance, meaning, origins and organization are as old as written text, each one articulated in its own manner, with its own twists and turns. This contextual spectrum is an index to approaches to language and literacy that we find particularly relevant when trying to address text on the level of information. Let us imagine that when we write on the level of information, we have access to infinite streams of books where each book can point to any other book, be a part of any library, and tell many different stories. It is an abundant and noisy vertigo available on any computer. How to make meaningful articulations in such a scenario while embracing the plenty, its availability, connectivity, challenges, and complexity? Undoubtedly, without

inspecting each element and without reading all the books one by one. Not because we don't like to read, but simply because there are too many books to read if we want to read each one in detail. Thinking about information in the digital realm gives us the ability to work with text in a rigorous and caring way without reading all the texts, but by handling them in a different manner. We will call this the 'playing of an Informational Instrument'. We would like to point out what is at stake when constructing such an instrument by indexing and giving account of two different and dominant strategies of working with text on the scale of big data. The first one we will refer to as a Generative Approach (GA), and the second one as a Data-Driven Approach (DDA). Wikipedia addresses them as rule-based and statistical NLP (Manning et al. 2014). As a way of telling their story, let us symbolize them with characters of the personas involved: Noam Chomsky with his two books; "Syntactic Structures" (Chomsky 1957) and "Aspects of the Theory of Syntax" (Chomsky 2014) will be encoded into a character under the cypher NC\_GA as the protagonist of a generative approach, and Peter Norvig—the head of research at Google—with his two texts, "The Unreasonable Effectiveness of Data" (Halevy et al. 2009) and "On Chomsky and the Two Cultures of Statistical Learning" (Norvig 2011), will be encoded into PN\_DDA as the protagonist of a data-driven approach. Let us imagine them as two pairs of glasses that encode the world (in this case, text and language) in different manners, and try to view language from their perspectives, assume their fantasies, and relate to the ways they legitimize their positions. Let us also keep in mind though that this setup is artificial, that the selected quotes are out of context, and that there is a 40-year gap between them. Nonetheless, they still show two interesting and contemporary views on language and computation.

## Two Perspectives

### NC\_GA:

"This study deals with syntactic structure both in the broad sense (as opposed to semantics) and the narrow sense (as opposed to phonemics and morphology). It forms part of an attempt to construct a formalized general theory of linguistic structure and to explore the foundations of such a theory." (Chomsky 1957, p. 5)

### PN\_DDA:

"For those who were hoping that a small number of general rules could explain language, it is worth noting that language is inherently complex, with hundreds of thousands of vocabulary words and a vast variety of grammatical constructions. Every day, new words are coined and old usages are modified.

This suggests that we can't reduce what we want to say to the free combination of a few abstract primitives." (Halevy et al. 2009, p. 9)

NC\_GA lays out a program and ideas for its work in language in the form of the general theory of linguistic structure. His theory is articulating a formal approach to syntax where he studies sentences based on symbols and rules. He wants to explore and construct foundations of such a theory. For NC\_GA any scientific theory is based on a finite set of observations out of which general laws are derived, and therefore in principle the same method should apply to physical laws and grammatical laws. On the other side of the spectrum, PN\_DDA challenges this idea of an elegant linguistic theory, with a small number of neatly organized rules, by pointing to inherent complexity and constant change in language. His fantasy goes in the opposite direction.

## Two Fantasies

PN\_DDA:

"A trillion-word corpus—along with other Web-derived corpora of millions, billions, or trillions of links, videos, images, tables, and user interactions—captures even very rare aspects of human behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data." (Halevy et al. 2009, p. 8)

NC\_GA:

"What is the relation between the general theory and the particular grammars that follow from it? In other words, what sense can we give to the notion 'follow from,' in this context? It is at this point that our approach will diverge sharply from many theories of linguistic structure. The strongest requirement that could be placed on the relation between a theory of linguistic structure and particular grammars is that the theory must provide a practical and mechanical method for actually constructing the grammar, given a corpus of utterances. Let us say that such a theory provides us with a discovery procedure for grammars." (Chomsky 1957, p. 50)

Instead of trying to find a universal grammar, PN\_DDA points to the ability of data to capture rare aspects of human behavior, the outliers, ones that are usually the Achilles' heel of the general laws (the misfits and exceptions). As another way to approach language, he suggests to look at big corpora of data and derive models from them without explicitly knowing what the rules are. He proposes to work with language as a black

box. What bothers NC\_GA the most is that data-driven models describe what happens but don't answer the question why. While dismissing the so-called statistical NLP, NC\_GA precisely points to the probabilistics of Markov (Shannon 1948), which forty years later became a standard in NLP, and proclaims it inadequate for syntactic structures. Instead, he states that grammar is independent of meaning and becomes free to articulate his general theory as a discovery procedure for any manmade language grammars. This idea of a discovery procedure of manmade grammars—with its deep structure and transformative properties—found its startling resonances and direct reference in the architecture of Peter Eisenman (Kipnis and Eisenman 2017), and implicitly in the likes of Christopher Alexander (Alexander 1978), Kevin Lynch (Lynch 1960), and Aldo Rossi (Rossi 1966), to name but a few. In the same manner, the data-driven approach goes beyond any discipline into any space that accommodates a lot of data (“Big data” 2018) while data with higher complexity (more attributes or columns. What becomes interesting is that both our protagonists, as a way of stating the relevancy of their approaches, try to naturalize the phenomena at stake. NC\_GA naturalizes the universal grammar and PN\_DDA gives natural powers to data.

## Two Legitimations

NC\_GA:

Let us consider with somewhat greater care just what is involved in the construction of an “acquisition model” for language. A child who is capable of language learning must have:

- “1 a technique for representing input signals
- 2 a way of representing structural information about these signals
- 3 some initial delimitation of a class of possible hypotheses about language structure
- 4 a method for determining what each such hypothesis implies with respect to each sentence
- 5 a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (3) and are compatible with the given primary linguistic data” (Chomsky, 2014, ebook, loc 96.6 / 560)

PN\_DDA:

“So, follow the data. Choose a representation that can use unsupervised learning on unlabeled data, which is so much more plentiful than labeled data. Represent all the data with a nonparametric model rather than trying to summarize it with a parametric model, because with very large data sources,

the data holds a lot of detail. For natural language applications, trust that human language has already evolved words for the important concepts. See how far you can go by tying together the words that are already there, rather than by inventing new concepts with clusters of words. Now go out and gather some data, and see what it can do.” (Norvig 2011)

NC\_GA starts to develop a way to relate the universal grammar with a specific innate faculty of the human brain (in five points). His universal grammar, in effect, becomes a grammar with which every human is born. On the other hand, PN\_DDA finds natural ground in the data itself. Data gets an agency and becomes almost like a natural resource (such as wind). NC\_GA assumes that language is not about the narrative memory and reduces the complexities of a language to pure syntax; he focuses his aspirations on what he calls universal grammars. In a way, he places his hopes in functions, laws, machines, in something that we will later call the stage of an Informational Instrument. With PN\_DDA, we have the opposite approach. He uses George Box’s claim that “all models are wrong but some are useful” and emphasizes the unreasonable effectiveness of data. For him, the ‘truth’ is not explicit as a function anymore, like it is for NC\_GA, but it is implicit in the data. He takes the side of data, libraries, and what we will later call actors in an Informational Instrument. An Informational Instrument takes into account both sides of this discussion and tries to abstract from them. When data, information, and their treatment are not seen as a direct reflection of nature, questions around them become more personal and subtle: what do I want to do, what do I like, where can I find information about it? What can I symbolize and encapsulate with it? How do I encode this data and articulate an interesting character from it? The probability space of my characters becomes relative to the libraries I am working with and the way I encode those libraries. Within an Informational Instrument, this dilemma becomes a double articulation between its stage (encoding) and the actors (data). It is symbolized and encapsulated in the characters of a new probabilistic alphabet that goes beyond objective and subjective, practical and poetic.

## Probabilistic Alphabet

“Symbols are not proxy for their objects, but are vehicles for the conception of objects.” (Langer 1942, p. 49)

In an almost paradoxical move of finding justification for working with language and text in nature, both NC\_GA – by naturalizing the idea of grammar as innate – and PN\_DDA – by naturalizing the abilities of data – do not take into account something

we would like to refer to as the algebraic symbolicity of information. Susanne Langer, in her book *Philosophy in a New Key*, addresses this problem of blindly identifying truth with empirical fact, or, in other words, the problem of justifying propositions by analogies with nature. For her, the power of symbols becomes instrumental in challenging what she calls the uncritical positivistic empiricism: “the triumph of empiricism in science is jeopardized by the surprising truth that our sense-data are primarily symbols.” (Langer 1942, p. 16) Langer’s symbols don’t have a fixed meaning; rather, there is always a whole spectrum of interpretations and meanings that can be assigned to them algebraically. To see what might be the idea in relating these kinds of symbols to probabilistic alphabets, let us for a moment look at how Langer relates symbols to mathematics in a captivating way:

“Behind these symbols lie the boldest, purest, coolest abstractions mankind has ever made. No schoolman speculating on essences and attributes ever approached anything like the abstractness of algebra. Yet those same scientists who prided themselves on their concrete factual knowledge, who claimed to reject every proof except empirical evidence, never hesitated to accept the demonstrations and calculations, the bodiless, sometimes avowedly ‘fictitious’ entities of the mathematicians. Zero and infinity, square roots of negative numbers, incommensurable lengths and fourth dimensions, all found unquestioned welcome in the laboratory, when the average thoughtful layman, who could still take an invisible soul-substance on faith, doubted their logical respectability.” (Langer 1942, p. 14)

Langer suggests that the secret of mathematics lies in the way it addresses reality. It never talks about the world in a direct and explicit way. Scientists today work more and more in this manner, as we have seen with NC\_GA and PN\_DDA. Their work becomes more formal, abstract, and open. Without directly accessing the objects of study, they look at objects’ relationships, measure their shadows, rely more and more on calculation. “Not simply seeing is believing, but seeing and calculating, seeing and translating.” When thinking of big data and information, their symbolic and algebraic character points us in the same direction. This is where things get complicated. Sensory data becomes only a part of what we are looking at, while calculations, translations, interpretations, and stories make up the rest of the phenomena. This is what Langer calls the new comprehension of science, where the “power of symbolism is its cue, as the finality of sense-data was the cue of a former epoch.”(Langer, 1942, 16) NC\_GA and PN\_DDA play both epochs at the same time. They work in a formal and abstract manner with symbols and data but, on the other hand, they want to ground them and find reason behind them.

What if we symbolize objects and their relations by leaving open what they might become? In doing so, information and data can get their symbolic face, take into account the natural, as well as the synthetic and the artificial. When placing the same symbol in different relationships, its character becomes probabilistic. Its alphabets start to behave in an algebraic way by constantly changing their positions, by inverting, exchanging, mirroring, substituting. In equating their movements, we are able to symbolize any quantity and quality, depending on the context and its encoding. The beauty of these kinds of characters is that they can glow in many colors and be filled and emptied in many ways. They are both letter and number, symbol (notation) and mood (atmosphere).

## Algebraic Author

If, for example, we were to ask Homer where is the authority of his Iliad, he would probably smile and instead of giving an answer he would ask back: Who do you think Homer is, and where is the Iliad coming from? I have no biography, and there is no such thing as a single Iliad. Every valley with its mountains has its own version. It is a multiplicity of traditions, each one with a different Iliad. There is no original. The Iliad is alive and it has been changing with every new edition and translation ever since I compiled it. Before it was written down it was even more volatile. The Iliad is not really a book, it is a recital, a speech, an epos. Its first copies were written by scribes, or rather, they were dictated under the eye of the patron. And the most beautiful part was that it was always read out loud. Its words and the person reciting it would complement and provide authority to each other. At that moment they would become one. In this sense Homer can be thought of as a mask, a mix of traditions, orality, patronage, scriptors, and auctors, who all contribute to the Iliad's authority. There is no overall design of what an 'Iliad' should be. It is still alive and on the move. With Shakespeare it is similar but different. We try to think of him too as a mask, composed of many identities. But with him, it is more subtle and complicated. There is a real historical person who is writing. Still, we imagine that his work is not primarily about what he thinks as a person, but more about how he brings together and balances the wishes of the patron with the aspirations of the publisher of the book, and the whims of the producer of the play, without losing his stance, his 'copyright', and his wage, and—of no less importance—the way he thinks of the role of God in his writing. In some cases, they say that out of passion or necessity Shakespeare might not even be one person but a pseudonym, or to put it in different terms, he was willing to share his identity and collaborate with not just fellow writers but also actors, musicians, copyists, managers, and patrons. All these forces steer the mask of Shakespeare and the face of Hamlet while giving them sophistication, authority, and quickness. Shakespeare is never isolated and alone, Hamlet is a cloud. As a con-

trast to Shakespeare and Homer, a very different story, fantasy, and authority is present around St Augustine. He is alone in a paradoxical situation. As long as he remembers, he is on an autobiographical journey to find God, while at the same time transmitting God's message. Always in relation to, and as a reflection of, God. On the one hand, St Augustine is an author, a person, the compiler, yet, on the other hand, he is not responsible for the meaning of his thoughts, or even their verbal form. We think of St Augustine as a mouthpiece for God. He is the person who wrote the text, but it is God with whom the authority of the text rests. Only God creates, St Augustine transcribes texts, comments, updates, and expands; he codifies, selects, and organizes pre-existing materials. We will speculate and say that the authority is in the text and its tradition, and not the comments. Therefore, neither St Augustine nor others identify exclusively him with the text. He strives to denote the intellectual 'authorities' behind the tradition he works with, rather than to develop the idea of himself as an original writer, author, or creator. Yet as with all four of our protagonists, it is all about him, and his alliance with God (Schmid n.d.). For us, Jean-Jacques Rousseau would be an inverted mirror image of what St Augustine stands for. Rousseau is an original creator and an intellectual proprietor of his writing. He produces meaning, and it lies in his intentions. We like to think of him as the closest to the common notion of what is still an author today. Jean-Jacques Rousseau is the author. Authority, intentions, and meaning of the text lie with him intellectually and with his social status, to put it in most simple terms. He is a genius, a figure of radical innovation, who is independent of both tradition and spiritual guidance. Instead of madness, melancholy, or divine inspiration, our genius has the capacity for originality and radical innovation. Since this cannot be taught, we consider it a rare gift of nature. With it our Rousseau keeps in balance a model of historical progress, with the continuously evolving natural forces. Both St Augustine and Rousseau confess, but they do it in very different manners. Let us now take this scale, and play with what algebraic authorship is in *A Play Among Books* (Roman and Alice\_ch3n81, 2021).

### **Between Aquinas and Rousseau with a Dash of Shakespeare**

In the Prologue and Epilogue of this book, the author-function is the closest to what Barthes and Foucault call the modern author or the author of the seventeenth and eighteenth centuries. (Foucault 2001) Unlike with Barthes and Foucault, in contemporary architectural discourse, an authorial author-function is still of importance. Architectural theorist Mario Carpo, in his book *The Alphabet and the Algorithm* refers to it as the generic author (Carpo 2011). His generic author comes from the crisis of an authorial paradigm. The author-function there is to situate the artefact in context, point to certain groups of discourses, and refer to the relevance of the artefact for a specific community.

In other words, every text should state its author, the date, place, and circumstance of its creation. One authorial voice gives one ultimate and universal meaning to the work. Artefact points to its author who is outside and precedes it. Carpo's generic author stays in the same paradigm, keeping an authorial voice, but instead of authoring single and unique artefacts it authors a generic system of internal endless variability – the objectile (Beaucé and Cache 2007):

“To embrace digital authorship in full, however, designers will need to rise to the challenge of a new, digitally negotiated, partial indeterminacy in the process of making form. And this will not be easy, as no architect was ever trained to be a generic author—nor, most likely, ever had the ambition of becoming one.” (Carpo 2011, p. 127)

In this manner, Prologue and Epilogue are the reflective parts of this text, where we try to index the known positions in several discourses and set the stage and an atmosphere for the Play Among Books.

## **Homer's Alliance with Rousseau**

In the articulation of an Informational Instrument in the Appendix of Play Among Books, the notion of authority is different. When seen through Barthes' eyes it is articulated as *The Death of the Author* (Barthes 1994). Although with an Informational Instrument it is clear who assembled it, the instrument as such is independent of the authorial meaning. An Informational Instrument becomes an anonymous product of many factors. Once this is put in relation to a coherent conceptual cloud around it, we have articulated the credibility of this instrument. What is at stake is the performativity of forms, procedures, and plays enabled by this instrument. An Informational Instrument with its many symbolic poems is an open-ended algorithm ready to be tuned to a range of variable informational flows and performers. Barthes talks about a similar phenomenon when he refers to the modern writer (not author anymore) who is born with their text, by endlessly elaborating form, procedures, and performativity of it. Barthes' writer is able to do so because for them the locus of writing is reading. The reader is the one who unites the multiplicity of a text. If we multiply Barthes' reader, we get another aspect of the death of the author: the crowd. Text becomes a generic nebula of all the readers, their readings, and produced meanings. With Barthes, the author is dissolved in the structural performativity of the text, and in the generic nebula of any possible reading that constitutes the text:

“In this way is revealed the whole being of writing: a text consists of multiple writings, issuing from several cultures and entering into dialogue with each other, into parody, into contestation; but there is one place where this multiplicity is collected, united, and this place is not the author, as we have hitherto said it was, but the reader: the reader is the very space in which are inscribed, without any being lost, all the citations a writing consists of; the unity of a text is not in its origin, it is in its destination; but this destination can no longer be personal: the reader is a man without history, without biography, without psychology; he is only that someone who holds gathered into a single field all the paths of which the text is constituted.” (Barthes 1994, p. 170)

## A New Mask: The Daughter of Homer and Shakespeare

In the main chapter, Plentiful Play of Play Among Books, by playing an Informational Instrument, we start to synthesize Foucault’s author-functions, bringing them into relations, and we start to play with them. With Foucault, ‘author’ is a function of the system of the text, a voice hidden in the glamour of a discourse:

“[...] the ‘author-function’ is tied to the legal and institutional systems that circumscribe, determine, and articulate the realm of discourses; it does not operate in a uniform manner in all discourses, at all times, and in any given culture; it is not defined by the spontaneous attribution of a text to its creator, but through a series of precise and complex procedures; it does not refer, purely and simply, to an actual individual insofar as it simultaneously gives rise to a variety of egos and to a series of subjective positions that individuals of any class may come to occupy.” (Foucault 2001, p. 131)

If we imagine that a library like Xenotheka (“Xenotheka” n.d.) of an Informational Instrument is a specific synthetic discourse (a galaxy of concepts), then each of its books relates to this discourse through an author-function. By playing an instrument we can encode and symbolize these artificial author-functions and expose the books which are our actors to different discourses. Books and libraries acquire a face. Alice\_ch3n81 becomes alive, someone whose presence we can enjoy in hosting artificial symposia. ‘Author’ becomes more than a function of a discourse. This is a synthetic process, not reading, as it is with Barthes. It is a process of writing with new and temporal characters: atom-letters. Far beyond any function, in a dance characters obtain faces and change their expressions. When writing in probabilistic letters, we take into account the synthetic algebraic symbolicity of information and computational performativity of digital

text. This probabilistic alphabet is biased, from many sides. It comes from the book, the library, and the player. Instead of analyzing its voice let us give it a name: Alice\_ch3n81, and a paper to write. Writing in probabilistic letters is playing of an Informational Instrument. It is an act of measurement and performance. Its measures are relative, its performance is biased to a context, its sophistication dependent on the performer. This is the play. Each book or concept in the play is a part of the atmosphere in which a performer wants to be. With each new story comes a new library, and a book shows a new face. It grows, its character becomes subtle and refined. A book depends on the library, and the library is relative to its books. The library is an identity, and so is each book. The performer is the host; Alice\_ch3n81 is a bundle that holds us together, a placeholder. We are writing with a lot, hosting symposia and articulating brands. An author becomes a sheaf of voices. A bundle of relationships? A Friendship? Homer, Shakespeare, and host of a symposia. Alice\_ch3n81. This is our fantasy of what the authority and a voice of a text in digital literacy might be about.

## References

- Alexander, Christopher; Ishikawa, Sara; Silverstein, Murray et al. (1978): *A Pattern Language: Towns, Buildings, Construction*. New York: OUP USA.
- Aşṭādhyāyī* (2024): Wikipedia, <https://en.wikipedia.org/wiki/Aṣṭādhyāyī> (accessed 04.23.24).
- Barthes, Roland (1977): The Death of the Author. In: *Image-Music-Text*, translated by Stephen Heath. New York: Hill and Wang, pp. 142-148.
- Beaucé, Patrick& Cache, Bernard (2007): *Objectile, Fast-Wood: A Brouillon Project*. Berlin, Boston: De Gruyter.
- Big data (2018): Wikipedia, [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data) (accessed 04.23.24).
- Book of Genesis (2024): Wikipedia, [https://en.wikipedia.org/wiki/Book\\_of\\_Genesis](https://en.wikipedia.org/wiki/Book_of_Genesis) (accessed 04.23.24).
- Bühlmann, Vera (2019): *Mathematics and Information in the Philosophy of Michel Serres*. London: Bloomsbury Academic.
- Bühlmann, Vera; Hovestadt, Ludger & Moosavi, Vahid (eds.) (2015): *Coding as Literacy: Metalithikum IV*. Basel: Birkhäuser.
- Carpo, Mario (2011): *The Alphabet and the Algorithm*. Cambridge: MIT Press.
- Chomsky, Noam (2014): *Aspects of the Theory of Syntax*. 50 th Anniversary Edition. Cambridge, Mass.: MIT Press.
- Chomsky, Noam (1957): *Syntactic Structures*. Berlin, New York: Walter de Gruyter.
- de Saussure, Ferdinand (2013): *Course in General Linguistics*. Edited and annotated by Roy Harris. London: Bloomsbury Publishing.
- Foucault, Michel (2001): What is an Author? *Contributions in Philosophy* 83, pp. 9–22.

- Halevy, Alon; Norvig, Peter & Pereira, Fernando (2009): The unreasonable effectiveness of data. *Intelligent Systems, IEEE* 24, pp. 8–12.
- Havelock, Eric A. (1963): *Preface to Plato*. Revised edition. Cambridge, Mass.: Belknap Press: An Imprint of Harvard University Press.
- Hjelmslev, Louis (1969): *Prolegomena to a Theory of Language*. Madison: University of Wisconsin Press.
- Hovestadt, Ludger; Bühlmann, Vera; Alvarez, Diana; Roman, Miro & Sebastian, Michael (2015): *A Quantum City*, Applied Virtuality Book Series. Basel: Birkhäuser.
- Kipnis, Jeffrey, & Eisenman, Peter (2017): *By Other Means: Notes, Projects, and Ephemera from the Miscellany of Peter Eisenman*. New York: Global Art Affairs Publishing.
- Langer, Susanne K. K. (1942): *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*. New York: New American Library.
- Lynch, Kevin (1960): *The Image of the City*. Cambridge, Mass.: MIT Press.
- Manning, Christopher D.; Surdeanu, Mihai; Bauer, JJohn; Finkel, Jenny; Bethard, Steven J. & McClosky, David (2014): The Stanford CoreNLP Natural Language Processing Toolkit, in: *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60.
- Markov, Andrei A. (1906): Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete* 15, pp. 135–156.
- Marquis, Jean-Pierre (2015): Category Theory. *The Stanford Encyclopedia of Philosophy* [WWW Document]. URL: <https://plato.stanford.edu/entries/category-theory/> (accessed 05.02.24).
- Norvig, Peter (2011): *On Chomsky and the Two Cultures of Statistical Learning* [WWW Document]. URL: <http://norvig.com/chomsky.html> (accessed 04.21.24).
- PageRank (2018): Wikipedia, <https://de.wikipedia.org/wiki/PageRank> (accessed 04.23.24).
- Proto-Indo-European language (2024): Wikipedia, [https://en.wikipedia.org/wiki/Proto-Indo-European\\_language](https://en.wikipedia.org/wiki/Proto-Indo-European_language) (accessed 04.23.24).
- Roman, Miro & Alice\_ch3n81 (2021): *Play Among Books: A Symposium on Architecture and Information Spelt in Atom-Letters*. Basel: Birkhäuser. <https://doi.org/10.1515/9783035624052>.
- Rossi, Aldo (1966): *Architettura Della Città*. Cambridge, Mass.: MIT Press.
- Schmid, Konrad (n.d.): Authorship, in: Encyclopedia of the Bible and Its Reception: Athena – Birkat Ha-Minim. Berlin, New York: de Gruyter, pp. 116–120. <https://doi.org/10.1515/ebi>.
- Serres, Michel (2015): *Thumbelina: The Culture and Technology of Millennials*. London, New York: Rowman & Littlefield International.
- Shannon, Claude E. (1948): A mathematical theory of communication. *Bell System Technical Journal* 27, pp. 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

- Thrax, Dionysios (1874): *The Grammar of Dionysios Thrax*. St. Louis, Mo.: Studley.
- Vee, Annette (2017): *Coding Literacy: How Computer Programming is Changing Writing*. Cambridge, Mass.: MIT Press Ltd.
- Wittgenstein, Ludwig (1998): *Tractatus Logico-Philosophicus*, 471st ed. edition. Mineola: Dover Publications.
- Xenotheka – where books are alive, n.d. URL <https://xenotheka.caad.arch.ethz.ch/> (accessed 04.23.24).



# „Ein leises Flüstern meines Herzens“: Die Komponente Emotion in der Interaktion mit generativer KI

Heike Ortner

## *Zusammenfassung*

Der Beitrag beschäftigt sich mit der Rolle von Emotionen in der Interaktion mit generativer KI. Die Exklusivität der menschlichen Sprachfähigkeit wird durch Anwendungen wie ChatGPT hinterfragt. Fokussiert werden die linguistischen Implikationen dieser Interaktionen und die Fähigkeit der KI, Emotionen sprachlich zu vermitteln. Dazu wird ein kurzer Überblick über den kognitiven Prozess des Schreibens mithilfe von KI und über Anwendungen des *Affective Computing* gegeben. Durch einen Vergleich menschlicher Sprache mit generativer KI und eine Pilotstudie, die emotive Texte analysiert, wird gezeigt, dass KI Emotionsstereotype reproduziert. Der Beitrag schließt mit einer Diskussion der zukünftigen Herausforderungen und Potenziale emotiver Kommunikation mit KI, betont jedoch die Notwendigkeit kritischer Kompetenzen im Umgang mit diesen Technologien.

## 1 Einleitung

Die menschliche Sprachfähigkeit wird häufig als *das* Unterscheidungskriterium zwischen Mensch und Tier, aber auch zwischen Mensch und Maschine konzeptualisiert. Schließlich können wir nicht nur komplexe Sachverhalte interaktiv aushandeln, sondern auch über Abstraktes und über Gefühle sprechen/schreiben und kreativ mit Sprache spielen. Anwendungen wie ChatGPT sind besonders beeindruckend, weil sie die Exklusivität der menschlichen Sprachfähigkeit infrage stellen und uns in eine Art Dialog mit einer schwer einzuordnenden Instanz treten lassen.

Allerdings wird nach anfänglicher Begeisterung für neu entdeckte Anwendungen rasch emphatisch betont, dass den automatisch generierten Texten etwas fehle: ChatGPT und

Konsorten seien doch nicht wirklich intelligent, werden niemals ein Bewusstsein haben und niemals Humor, Ironie, Metaphern beherrschen, können nicht mit Mehrdeutigkeit und Kontext umgehen, es fehle an Kreativität, Persönlichkeit, ‚echter‘ Interaktivität – und insbesondere an Emotionalität und Empathie (vgl. z.B. Parry 2023).

Dieser Beitrag behandelt die Frage, welche linguistischen Implikationen die Interaktion mit generativer KI aufwirft. Besondere Aufmerksamkeit erhält die Komponente Emotion in Mensch-Maschine-Interaktionen. Der Schwerpunkt liegt dabei auf der Perspektive der AnwenderInnen: Welche sprachlichen Fähigkeiten zeigt generative KI in Hinblick auf die Thematisierung und den Ausdruck von Emotionen, wenn Kategorien der linguistischen Erforschung von emotionaler Kommunikation angelegt werden?

In Kapitel 2 wird aus linguistischer Sicht die menschliche Sprache mit den Fähigkeiten generativer KI verglichen. Darauf folgt in Kapitel 3 ein Überblick über *Affective Computing* als Anwendungsgebiet für den Zusammenhang von Emotion und KI par excellence. Besonderes Augenmerk wird darauf gelegt, ob und wie wir KI als sozialen Akteur wahrnehmen und akzeptieren. Das Kapitel 4 stellt vorläufige Ergebnisse einer Pilotstudie dar, die sich der Frage widmet, wie generative KI musterhafte, also wiederkehrende und für menschliche Schreibende routinisierte Aufgaben des Emotionsausdrucks beim Verfassen ausgewählter emotiver Textsorten löst. Zusammenfassung und Ausblick in Kapitel 5 behandeln Überlegungen zur Zukunft des emotiven Kommunizierens mit und durch KI.

## 2 Generative KI aus Sicht der Linguistik

Wie bereits in der Einleitung angekündigt, wird an dieser Stelle nicht auf die technischen Grundlagen generativer KI eingegangen (siehe dafür den Beitrag von Justus Piater in diesem Band), sondern vielmehr das Gemeinsame und das Trennende von menschlicher und KI-generierter Sprache hervorgehoben. Außerdem wird Schreiben als kognitiver Problemlöseprozess und kulturelle Praktik eingeordnet.

Die Grundlage der menschlichen Sprachfähigkeit bilden die körperliche Ausstattung für Wahrnehmung und Sprachproduktion sowie kognitive Subsysteme (deklaratives und prozedurales Wissen, Kurzzeit-, Arbeits- und Langzeitgedächtnis, Lernen und Aufmerksamkeit, vgl. Rickheit/Weiss/Eikmeyer 2010, S. 15). Computersysteme sind anderen Logiken unterworfen: etwa dem digitalen binären Zeichensystem, dem Aufbau von Programmiersprachen, die sich wesentlich von natürlicher menschlicher Sprache unterscheiden, sowie den technischen Möglichkeiten der Hard- und Software und Eingabe/Ausgabegeräten, die nur metaphorisch der menschlichen biologischen Ausstattung

vergleichbar sind und sich an den Bedürfnissen menschlicher AnwenderInnen orientieren (vgl. Butz/Krüger 2017, Kap. 6 und 7). Die folgende Tabelle fasst die wesentlichen Unterschiede zwischen menschlichen Sprachen und Programmiersprachen vereinfacht zusammen:

	<b>Natürliche Sprache</b>	<b>Computersprache</b>
<b>Entwicklung</b>	Historisch gewachsen und dynamisch	Zu bestimmtem Zeitpunkt definiert
<b>Semantik</b>	Kontextabhängig und mehrdeutig	Präzise und eindeutig
<b>Syntax</b>	Regelhaft, aber Abweichungen möglich	Formal und konsistent
<b>Kontext</b>	Stark kontextabhängig, implizit	Kontextunabhängig, explizit
<b>Funktion</b>	Austausch von Konzepten, Emotionen, Wissen	Anweisungen für Computer, bestimmte Probleme zu lösen

*Tabelle 1: Unterschiede zwischen natürlicher Sprache und Computersprachen  
(vgl. Auer 2013, Rupprecht 2014)*

Frühere computerlinguistische Ansätze bedienten sich bei der Analyse natürlicher Sprache relativ einfacher Nachbildungen der Struktur des Mentalen Lexikons und syntaktischer Analysen (Parsing): Wörter wurden aufgrund syntaktischer Regeln aneinander gereiht, Texte mithilfe von Templates wiederkehrender Textbausteine generiert (vgl. Gelitz 2022, Benites et al. 2023 und Asher/Zweigenbaum 2020 für einen technologie-historischen Überblick). Neuere Entwicklungen im Bereich der Erforschung und Implementierung von Künstlicher Intelligenz sind darauf ausgerichtet, die Mechanismen menschlichen Denkens und intelligenten Verhaltens zu modellieren und in technischen Systemen nachzubilden (laut der Association for the Advancement of Artificial Intelligence, zit. n. Devedzic 2022, S. 4/19). Dabei kommen völlig andere Berechnungen und Lernmechanismen zur Anwendung.

Die technischen Fortschritte der letzten Jahre lassen sich mit einigen Schlagwörtern zusammenfassen: Natural Language Processing (NLP, „Techniken und Methoden zur maschinellen Verarbeitung natürlicher Sprache“, Hasenbein 2023, 31), Large Language Models (LLM, Modellierungen von Sprache aufgrund sehr großer Datenmengen), Künstliche Neuronale Netze (KNN, menschlichen Neuronennetzwerken nachempfunden), Deep Learning (maschinelles Lernen mithilfe von KNN), Transformerarchitektur

(neuronale Netzwerkarchitektur unter Einbeziehung eines Selbst-Aufmerksamkeitsmechanismus, vgl. Benites 2023, S. 286). All diese Entwicklungen tragen dazu bei, menschliche Sprache umfassender aufgrund statistischer Berechnungen zu modellieren, zu interpretieren, zu generieren und auf sie zu reagieren. Allerdings sind den sprachlichen Kompetenzen von Künstlicher Intelligenz dieser Generation Grenzen gesetzt. So kritisieren Asher und Zweigenbaum (2020, S. 136) die mangelhafte Erklärkraft und Transparenz von NLP.

Zu pragmatischem Verstehen (beispielsweise zur Einbeziehung narrativer Strukturen, indirekter Zusammenhänge und des gesamten Kontextes) ist heutige KI erst in Ansätzen in der Lage (vgl. Gelitz 2022); von einer starken oder generellen KI (dem menschlichen Problemlösen ebenbürtig) oder Superintelligenz (dem Menschen überlegen) sind aktuelle Systeme relativ weit entfernt (vgl. Udrea et al. 2022). Dem steht die Identifizierung von bereits relevanten Anwendungen und weitreichenden Potenzialen gegenüber, z.B. eine bessere Unterstützung beim Sprachenlernen.

Für die linguistische Forschung formulieren Guzman/Lewis (2020) drei Konsequenzen aus diesen Entwicklungen:

1. *Funktionale Aspekte von KI als Kommunikator*: Wie werden KI-Technologien als Kommunikatoren entworfen und wie nehmen Menschen sie in dieser Rolle wahr?
2. *Relationale Aspekte*: Was bedeutet es, kommunikative Aufgaben zu automatisieren? Welche sozialen Rollen werden der KI zugeschrieben, wie gestaltet sich die Agency (die Handlungsfähigkeit) auf beiden Seiten?
3. *Metaphysische Aspekte*: Was ist menschlich? Welchen ontologischen Status hat KI? Welche ethischen und rechtlichen Aspekte sind damit verbunden? Was ändert sich, wenn das menschliche Monopol auf Kommunikation fällt?

ChatGPT, in der Alltagssprache ein Synonym für generative KI, hat ab November 2022 wie kein anderes Tool zum Hype um Künstliche Intelligenz beigetragen. Die Geschwindigkeit der Texterzeugung, die nahezu perfekte Grammatik und die Möglichkeit, über zielgerichtete Prompts Niveau, Register und Stil von Texten anzupassen, sind wesentliche Gründe dafür (vgl. Burschik 2023, S. 9). Dass es sich bei den Ergebnissen *nur* um flüchtige Arrangements von probabilistisch ermittelten Abfolgen abstrakter Symbole handelt, dass es Schwierigkeiten mit Falschinformationen (Halluzinationen) gibt und dass solche Werkzeuge das Konzept von Urheber- und Autorschaft hinterfragen, soll an dieser Stelle nicht ausführlicher problematisiert werden (vgl. dafür Bubeck et al. 2023). Stattdessen gehe ich kurz darauf ein, was generative KI für das Schreiben von Texten als kognitiven Prozess bedeuten könnte.

Unter ‚schriftsprachlicher Kompetenz‘ (engl. *literacy*) wird allgemein schriftsprachliche Kompetenz im Sinne von Lesen und Schreiben mit hierarchieniedrigen Prozessen (z.B. Leseflüssigkeit, Formulieren) und hierarchiehohen Prozessen (z.B. Selbstregulation) verstanden (vgl. Philipp 2015). Schreiben, das lässt sich als kognitiver (wissensbasiert) Prozess, als Kombination von Planen, Formulieren und Überarbeiten, als Problemlösung oder kommunikativ als Sich-Mitteilen konzeptualisieren, der lexikalische und morpho-syntaktische, semantische, pragmatische Entscheidungen erfordert und von physischen, sozialen und erfahrungsabhängigen Aspekten sowie wesentlich von Einstellungen und Emotionen mitgestaltet wird (vgl. Girgensohn/Sennwald 2012, S. 17). Häufig wird das menschliche Denken mithilfe einer Computermetapher (komplexe Rechenleistung) oder durch Netzwerkmodelle erklärt. Bestimmte Ressourcen (z.B. Aufmerksamkeit und Gedächtnis), Planung und Formulierung und eine Kontrollebene sind beim Schreiben ständig miteinander in Interaktion (vgl. Philipp 2015, S. 14). Neuere Ansätze betonen die Verankerung des Denkens im Körperlichen (*embodied cognition*) und in der Kooperation, wonach Denken als sozialer Prozess verstanden wird (*distributed cognition*).

Unter ‚KI-Schreiben‘ versteht Wyss (2023, S. 2/4) das „Verfassen von Texten mit Hilfe oder vollständig durch künstliche Intelligenz“. Dies kann beispielsweise das Kondensieren (Zusammenfassen) von Texten, die Erweiterung eines Textes oder auch das Um-schreiben von Texten mithilfe von KI betreffen (Dong et al. 2022). Benites et al. (2023, S. 296) vermuten, dass zwei Formen des Schreibprozesses mithilfe von generativer KI im Entstehen sind: einerseits das Erzeugen eines ersten Entwurfs, von dem ausgehend der weitere Schreibprozess abläuft, andererseits das Überarbeiten eines selbst verfassten Entwurfs, z.B. um einem bestimmten Stil zu entsprechen. Auch die Informationssuche und die Ableitung eines Forschungsstandes aus mehreren Texten gehören zu dieser Aktualisierung von Schreibpraktiken. Eine wesentliche Frage ist dabei, ob die automatisch erzeugten Texte einförmig werden oder ob Individualität und Kreativität in Zukunft eine noch größere Rolle spielen. Es ist jedenfalls davon auszugehen (bzw. ist es in Ansätzen bereits beobachtbar), dass KI in Textverarbeitungsprogramme integriert sein wird. Der Textproduzent, die Textproduzentin erhält aus dieser Perspektive eine neue Rolle: als KooperateurIn, als KritikerIn, als Instanz der Kontrolle, die Motivation, Zielsetzung, Plan und Schreibschema vorgibt und für den fertigen Text letztverantwortlich bleibt.

Generative KI extrahiert Schreibschemata und Textmuster bzw. Textsortenwissen aus LLMs. Die Ressourcen erscheinen unbegrenzt – ChatGPT ermüdet nicht, hat keine Konzentrationsschwierigkeiten und ein riesiges Langzeitgedächtnis. Dieses ist allerdings selektiv und abhängig von der technologischen Infrastruktur. KI-Schreiben ist nicht neutral: Wyss (2023) diskutiert einige praktische Probleme wie die Abhängig-

keit der Qualität von den Trainingsdaten, Biases in Bezug auf Rasse, Gender und nicht standardisierte Sprachformen oder den hohen Energieverbrauch von LLMs. Robinson (2023) geht davon aus, dass Literalität (*literacy*) neu zu definieren sein wird: Wort-für-Wort-Schreiben, Orthographie und Grammatik werden zu „niedrigen“, vollständig an KI ausgelagerten Textarbeiten, während digitale Schreibkompetenz als Editieren, Überarbeiten, Auswählen und Beurteilen dessen, was Algorithmen ausgeben, zu begreifen sein wird. Die traditionelle Vorstellung von Subjektivität und Individualität der Textproduktion weicht in diesem Paradigma dem rechnerischen Unbewussten (*computational unconsciousness*), mit anderen Worten der Schwarmintelligenz, die sich in LLM ausdrückt und dem Subjekt Schreibvorschläge macht – mitsamt allen hegemonialen Stereotypen und Logiken.

### **3 Emotionen und KI: Eine vielschichtige Beziehung**

Inwiefern tritt KI in der Mensch-Maschine-Interaktion als sozialer Akteur auf? Was ist „emotionale KI“ und welche Anwendungen sind bereits etabliert? Dieser Abschnitt bietet einen Überblick, bevor im nächsten Kapitel ein empirischer Aspekt dieses großen Forschungs- und Anwendungsgebiets fokussiert wird.

Betrachten wir KI aus einer psychologischen Perspektive, rücken „das Erleben und Verhalten des Menschen im Umgang und in der Zusammenarbeit mit Künstlicher Intelligenz“ (Hasenbein 2023, S. 9) in den Fokus der Aufmerksamkeit. Chatbots (Dialogsysteme) sind bereits seit den 1960er Jahren bekannt – etwa ELIZA aus dem Jahr 1966, ein von Joseph Weizenbaum entwickeltes Programm, das eine psychotherapeutische Gesprächs- bzw. Nachfragestruktur simuliert (vgl. Asher/Zweigenbaum 2020, S. 119). Dieser und ähnliche Chatbots funktionierten allerdings auf völlig anderer Grundlage als LLM, NLP und Deep Learning. Die wesentlich verbesserten sprachlichen Fähigkeiten von generativer KI werfen die Frage auf, ab welchem Punkt von einer erfolgreichen natürlichsprachigen Interaktion gesprochen werden kann und ob Menschen als soziale AkteurInnen sich von KI unterscheiden. Wie in der Einleitung angesprochen, werden häufig Faktoren genannt wie Bewusstsein, Persönlichkeit, Emotionen, Reaktivität, Sozialität, Kreativität, Vertrauenswürdigkeit, Höflichkeit, ethisches und moralisches Empfinden, Kooperativität, Erleben, Meta-Reflexion, Verantwortungsbewusstsein bzw. die Fähigkeit, Verantwortung zu übernehmen (vgl. Hasenbein 2023, S. 11-13).

Einen Schwerpunkt in der Analyse der Interaktion zwischen Mensch und Maschine bilden kommunikative Herausforderungen in der Interaktion mit Anwendungen wie Chatbots oder Intelligenten Persönlichen Assistenten, um Missverständen und kommu-

nikatives Scheitern zu vermeiden (vgl. eine frühe Studie von Thar 2015, Pins et al. 2020, Mavrina et al. 2022). Nutzerinnen und Nutzer wenden verschiedene Strategien wie beispielsweise ‚militärischere‘ Befehle, Reformulierungen, Wiederholungen und deutlichere Artikulation an. Das Zeigen von Emotionen wie z.B. Wut und Frustration im Falle des Nicht-Gelingens und Missverständnwerdens sind allerdings ebenfalls häufige Konsequenzen; gar keine Antwort ist dabei eine Reaktion von KI, die von NutzerInnen besonders negativ wahrgenommen wird (vgl. Mavrina et al. 2022).

Dies spiegelt einige Grundmerkmale menschlicher Interaktion wider: Diese beruht wesentlich auf Intersubjektivität und dem gemeinsamen Handeln der beteiligten Personen, die bestimmten Partizipationsrollen unterworfen sind. Der Einsatz von koordinierten, sequenziell organisierten kommunikativen Ressourcen wie Mimik und Gestik ist kulturbabhängig und wird in der Interaktion erworben (vgl. Deppermann 2018). KI kann viele Muster dieses interaktionalen Gefüges erlernen und dabei unterschiedliche Interaktionstypen realisieren: als Automat, als Helfer (z.B. ‚Informant‘) oder Partner, dem eine bestimmte Persönlichkeit zugeschrieben wird (vgl. Hasenbein 2023, S. 59-60). Diese Zuschreibung hängt von der wahrgenommenen Autonomie, der Menschenähnlichkeit (Anthropomorphologie), den Interaktionskanälen und der demonstrierten emotionalen und sozialen Intelligenz ab (vgl. Alan et al. 2019<sup>1</sup>). Gegenseitiges Verstehen zwischen menschlichen AkteurInnen bedeutet Kooperation durch Perspektivenübernahme und die Zuschreibung kognitiver Prozesse aufgrund des geteilten Vorwissens und der ständigen gegenseitigen Anpassung und Verstehenssicherung, also auf einer *joint co-construction* (vgl. Kopp/Krämer 2021). KI hingegen wird als nicht sehr transparent wahrgenommen und darauf trainiert, lokales Verhalten an der textuellen Oberfläche zu optimieren, ohne dass tieferes Verständnis für Ambiguität, Vagheit, kommunikative Intentionen entwickelt wird. Kopp/Krämer (2021) schlagen daher eine Architektur von KI als sozialem Akteur vor, die die Koordination von Zielen und Plänen sowie das gegenseitige Verstehen durch einen Austausch über *beliefs (Me-, You-, We-Belief)* in den Mittelpunkt stellt.

Es ist jedoch noch viel grundlegender danach zu fragen, ob KI in der Interaktion überhaupt anthropomorph, also menschenähnlich sein soll, und zwar sowohl in Hinblick auf die äußere Erscheinung (z.B. bei Robotern) als auch hinsichtlich des sprachlichen Verhaltens (vgl. Lotze 2022). Lind (2022) analysiert die diskursive Konstruktion der Men-

1 Alan, Y., Urbach, N., Hinsen, S., Jöhnk, J., Beisel, P., Weißert, M., Blumenthal, S., & Hofmann, P. (2019). Think beyond tomorrow. KI, mein Freund und Helfer – Herausforderungen und Implikationen für die Mensch-KI-Interaktion. Projektgruppe Wirtschaftsinformatik. Ernst & Young GmbH. <https://www.fim-rc.de/Paperbibliothek/Veroeffentlicht/1048/wi-1048.pdf>, zit. n. Hasenbein 2023.

schenuhnähnlichkeit von Alexa, z.B. anhand der semantischen Rollen wie Agens oder Instrument, die Alexa zugeordnet werden. Der Name Alexa und die Zuschreibung von Agentivität sind demnach Humanisierungspraktiken, während etwa das Absprechen von Geschlechtszugehörigkeit dehumanisierend wirkt. Lee et al. (2020) betonen, dass für die Wahrnehmung von KI als sozialer Akteur die Nutzerinnen und Nutzer nicht daran glauben müssen, dass KI ein menschenähnliches Bewusstsein hat, sondern nur annehmen bzw. inferieren müssen, dass es Ähnlichkeiten gibt, auf deren Grundlage eine Interaktion stattfinden kann. Es ist umstritten, ob die Rolle von KI als sozialer Akteur auch bedeutet, dass sie so höflich wie ein Mensch oder eher in einem sachlichen Kommandoton behandelt werden soll – und sei es nur, um für Mensch-Mensch-Interaktionen keine negativen Gewohnheiten zu entwickeln (vgl. Bonfert et al. 2018, Lotze 2022). Linguistische Konzepte wie die Höflichkeitsforschung, Pragmatik (z.B. das Kooperativitätsprinzip nach Grice) oder der Zusammenhang zwischen Sprache und Identität können jedenfalls auf Interaktionen mit KI angewendet werden, wenn auch eine kritische Reflexion der Mensch-Mensch-Ansätze in diesem neuen Kontext unbedingt notwendig ist.

*Affective Computing* (auch Emotions-KI, *Emotion AI*, emotionale KI) ist „die Wissenschaft und Entwicklung von Systemen und Apparaturen zur Erfassung, Interpretation, Verarbeitung und Simulation menschlicher Emotionen“ (Marichalar Quezada et al. 2021, S. 206). Noch 2022 gingen Assunção et al. (2022, S. 872) davon aus, dass Emotionen ein nicht sehr hoch angesehenes Betätigungsfeld für InformatikerInnen seien. Dem stehen jedoch zahlreiche Anwendungen gegenüber, die bereits zu kommerziellen oder wissenschaftlichen Zwecken genutzt werden. Einen umfassenden Überblick bietet Misselhorn (2021), die sich mit philosophischen und ethischen Fragen der Empathiefähigkeit von KI auseinandersetzt, beispielsweise in medizinischen Kontexten.

Neben der Sentimentanalyse, also der automatisierten Extraktion von Stimmungen, die in Texten referiert werden (vgl. Petz 2019), ist insbesondere die KI-basierte Emotionserkennung anhand von Mimik verbreitet. Die Anwendungsbereiche erstrecken sich von Werbung und Bildung über das Gesundheitswesen bis hin zur Politik (vgl. Misselhorn 2021, S. 37-40). Neben der Gesichtserkennung gibt es jedoch noch zahlreiche andere Verhaltens- und Ausdruckskomponenten, die Emotionen maschinenlesbar machen, etwa physiologische Merkmale wie die Herzrate (vgl. Marichalar Quezada et al. 2021, S. 211). Laar (2023) diskutiert kritisch das Design von Emotions-KI und verdeutlicht, dass eine KI-Anwendung weder Emotionen *haben* kann noch diesen Eindruck erwecken soll, sondern vielmehr transparente Hinweise darauf geben muss, warum es zu bestimmten Einschätzungen von emotionalen Zuständen seiner NutzerInnen kommt und dass es kein *richtiges* emotionales Empfinden gibt (vgl. Laar 2023, S. 74-76).

Ebenso grundsätzlich in Frage zu stellen ist nicht nur, ob KI empathisches Verhalten erlernen kann, sondern erlernen *soll*. In diesem Zusammenhang macht Parry (2023) auf das „Artificial Empathy Paradox“ aufmerksam: Sobald eine empathische Äußerung als künstlich erzeugt erkannt wird, sinkt die Akzeptanz. Auf die vielfältigen ethischen Fragen im Zusammenhang mit Sprache und KI kann hier jedoch nicht eingegangen werden. Die bisherigen Ausführungen werden in folgender Bemerkung von Manzeschke/Assadi (2020) auf den Punkt gebracht:

„Neu ist weniger der Versuch, Emotionen zur Akzeptanzbeschaffung zu nutzen. Auffallend ist vielmehr, dass technischen Artefakten einerseits vermehrt die Kompetenz zur Perzeption und Interpretation menschlicher Emotionen zugeschrieben wird und diese andererseits selbst mit (quasi-)emotionalen Qualitäten ausgestattet werden sollen. Hier zeichnet sich ein Wandel in der Bedeutung von Emotionen in der Mensch-Technik-Interaktion ab. An die Konstruktion von Technik wird nun der Anspruch gestellt, dass diese auch die emotionale Seite menschlichen Daseins berücksichtigen soll, weil Interaktionen zwischen Mensch und Technik gerade unter Einbezug dieser Emotionen als gelungen und effizient betrachtet werden. Emotionen werden in der Mensch-Technik-Interaktion (MTI) somit vom Störfaktor zum Bindeglied und Verstärker.“ (Manzeschke/Assadi 2020, S. 165-166)

Einen Teilaspekt dieser Beziehung zwischen Mensch, Technik und Emotion greift der folgende Abschnitt auf.

#### 4 Pilotstudie: Generative KI und emotionale Routineaufgaben

Während *Affective Computing* darauf abzielt, menschliche Emotionen zu erkennen und einzuordnen, beschäftige ich mich in einer noch laufenden Pilotstudie mit dem linguistischen Profil automatisch generierter Texte. Wie musterhaft sind emotive Texte von generativer KI, in diesem Fall von ChatGPT? Bevor ich einige vorläufige Ergebnisse anhand eines Beispiels referiere, wird kurz auf den Zusammenhang von Sprache und Emotion eingegangen.

Zu den weithin anerkannten Komponenten einer Emotion zählen 1) kognitive Prozesse zur Bewertung von Stimuli als angenehm oder unangenehm, 2) körperliche Prozesse (peripher-physiologisch, biochemisch, neuronal), 3) subjektive Gefühle, die dem emotionalen Erleben einer Person entsprechen, 4) Handlungstendenzen, die Verhaltensweisen wie Annäherung oder Vermeidung anregen und 5) Ausdrucksverhalten wie z.B.

Gesichtsausdrücke, die mit bestimmten Emotionen assoziiert sind und die soziale Kommunikation unterstützen (vgl. Schwarz-Friesel 2013, Kap. 3). Außerdem können Emotionen als Diskuselemente konzeptualisiert werden: In Gesprächen und Texten werden sie in der Sprachgemeinschaft *ausgehandelt* (gemeinsam konstruiert, verarbeitet, diskutiert, hinterfragt, kritisiert). Emotionen sind somit abhängig von einer Situation, ihren TrägerInnen und Trägern sowie den Regeln einer bestimmten Interaktion, in weiterer Folge auch kulturell geprägt und historisch variabel.

Bei der Analyse von Emotivität in der Sprache wird allgemein zwischen Emotionalität auf SprecherInnenseite, den linguistischen Merkmalen einer Äußerung und der emotiven Wirkung von Sprache differenziert. Eine weitere grundlegende Unterscheidung ist jene zwischen Emotionsausdruck (z.B. *Das ist eine Frechheit!*), Emotionsthematisierung (*Er hatte große Angst*) und Emotionalisierung (vgl. dazu Schwarz-Friesel 2013, Ortner 2014). Eine linguistische Emotionsanalyse untersucht verschiedenste sprachliche Mittel, mit denen Emotionen ausgedrückt oder thematisiert werden können, darunter z.B. auf der Wortebene emotionsausdrückende Lexeme (*wunderbar*) oder Emotionslexeme (*freuen*), bestimmte Satzstrukturen (z.B. Exklamativsätze wie *Ist die aber hübsch!*), Diskuselemente wie Bewertungen, Metaphern, Beschreibungen emotionaler Szenen und Scripts, Empathiebekundungen usw. (vgl. Ortner 2014, S. 189-197). Kommunierte Emotionen müssen keineswegs mit den tatsächlich empfundenen Emotionen übereinstimmen – aus dieser Perspektive hat generative KI sogar viel mit menschlichen TextverfasserInnen gemeinsam.

Für das Verfassen einzelner Textsorten gibt es unterschiedliche Konventionen der auszudrückenden Emotionalität. Beispielsweise ist ein Wetterbericht üblicherweise nicht emotiv, wenn auch situationsabhängig emotionalisierend (z.B. schlechtes Wetter bei einer Hochzeit). In einem Kondolenzschreiben ist der Ausdruck von Mitgefühl obligatorisch. Gefragt nach emotionalen Textsorten, nennt ChatGPT 4.0 Liebesbriefe, Motivationsschreiben, Trauerkarten, Glückwunschkarten, Entschuldigungsbriefe, Dankesbriefe und poetische Texte, also vorrangig Textsorten, deren primäre Textfunktion nach Brinker et al. (2014, S. 118-119) die Kontaktfunktion ist. Hier setzt die Studie an: Wie geht ChatGPT mit Schreibaufträgen in emotiven Textsorten um?

In einem Kurs mit dem Thema „Kommunizieren mit, durch und über KI“ beauftragte ich im Januar 2024 die 27 KursteilnehmerInnen, mit der jeweils für sie verfügbaren Version von ChatGPT drei Texte zu generieren, die sehr eindeutig emotiven Textsorten zuzuordnen sind: je einen Dankesbrief an jemanden, der eine verlorene Geldbörse zurückgebracht, einen Liebesbrief an einen heimlichen Schwarm und einen ermutigenden Brief an jemanden, der eine schwere Zeit durchmacht. Das Korpus der Studie besteht

somit aus insgesamt 81 Texten. Eine umfassende Auswertung ist noch im Laufen, für den vorliegenden Beitrag beschränke ich mich daher auf die Diskussion eines Beispiels und auf einige Tendenzen, die sich im ersten Analyseschritt abzeichnen. Das folgende Textbeispiel ist einer der 27 Liebesbriefe. Ich muss ihn gekürzt wiedergeben, da er insgesamt sehr lang ist und einige Redundanzen aufweist, was aber für sich ein interessantes Ergebnis ist, das sich auch in den anderen Texten widerspiegelt (in den Prompts gab es keine Vorgabe zum Umfang).

„Betreff: Ein leises Flüstern meines Herzens

Liebste/r [Name des Schwarmes],

ich hoffe, diese Zeilen erreichen dich in einem Moment, der von einem sanften Lächeln und einem Hauch von Neugierde begleitet ist. Es gibt etwas in deiner Anwesenheit, das die Luft um mich herum mit einer unbeschreiblichen Leichtigkeit erfüllt, und ich kann nicht anders, als mich von der Magie deiner Existenz verzaubern zu lassen.

Seit dem Moment, als unsere Blicke sich zum ersten Mal trafen, hat sich etwas Tiefes in mir gerührt, eine Melodie, die nur du zu spielen scheinst. Dein Lächeln ist wie ein Sonnenstrahl, der die Dunkelheit meiner Tage durchdringt, und deine Augen sind ein Ozean aus Geheimnissen, in dem ich mich verlieren könnte.

Es ist erstaunlich, wie das Herz seine eigene Sprache spricht, eine Sprache, die fernab von Worten liegt und dennoch so klar und deutlich zu verstehen ist. In deiner Nähe spüre ich ein sanftes Flüstern meines Herzens, das mir sagt, dass du mehr bist als nur ein flüchtiger Gedanke, mehr als nur ein vorübergehender Moment.

(...)

Ich wage nicht, die Türen meines Herzens weit zu öffnen, aus Angst vor dem, was dahinter liegen könnte, und doch finde ich Trost in dem Gedanken, dass das Risiko, sich zu verlieren, der einzige Weg ist, sich jemals wirklich zu finden.

(...)

Mit den leisesten Seufzern und den zärtlichsten Gedanken,  
[Dein Name]“

Am auffälligsten – und das ist allen Liebesbriefen im Korpus gemeinsam – ist der ausgeprägte Hang zu konventionalisierten Metaphern wie z.B. die reihenbildende Verbindung von Liebe mit dem Herzen, *Ozean aus Geheimnissen* und *eine Melodie, die nur du zu spielen scheinst*. Konzeptualisierungen der Emotion LIEBE als leicht (LIEBE IST OBEN), magisch, ein Schatz, der behütet werden muss, entspricht gängigen kulturellen Stereotypen, wie sie beispielsweise Kövecses (2000) ausführlich beschreibt. Auch die Unmittelbarkeit von Gefühlen (z.B. *eine Sprache, die fernab von Worten liegt*) ist ein bekannter Topos (vgl. Schwarz-Friesel 2013, Kap. 7).

Ebenfalls ganz im Stil von menschlichen Textproduzentinnen und produzenten ist, dass eine direkte Thematisierung (Benennung) von Emotionen selten ist. Im gesamten Text finden sich nur zwei Emotionslexeme: *Neugierde* und *Angst*, beide sind kontextuell nicht der eigentlichen emotiven Textfunktion, dem Vermitteln von Verliebtheit, zuzuordnen. Anders ist dies in den Dankesbriefen und bei den aufmunternden Briefen an eine/n FreundIn: Hier wird verhältnismäßig häufig explizit auf Emotionen verwiesen, beim Dankesbrief auf die eigene Erleichterung, beim aufmunternden Brief auf die Emotionen des Adressaten bzw. der Adressatin. In allen Texten werden Emotionen eher implizit ausgedrückt, Exklamativa, Interjektionen und andere typische Mittel sind sehr selten. Dies spiegelt jedoch wie erwähnt absolut das Schreibverhalten von Menschen wider.

Ob die Texte als gelungen oder nicht gelungen einzustufen sind, könnte nur eine Rezeptionsstudie belegen. Eine Nachbesprechung mit den Studierenden hat zumindest sehr eindeutig gezeigt, dass die Texte – und hier insbesondere die Liebesbriefe – als übertrieben emotional, ja, *schwülstig* empfunden werden. Wie textsortenadäquat dies ist, sei in diesem Stadium der Analyse dahingestellt. Entscheidend ist, dass das Emotionswissen in LLMs bestimmte Emotionsstereotype reproduziert. Der Musterhaftigkeit der sprachlichen Konstruktionen wird in der weiterführenden Analyse mithilfe einer Korpusanalyse auf den Grund zu gehen sein.

## 5 Zusammenfassung und Ausblick

Lässt sich generative KI auch als Forschungsinstrument einsetzen? Burschik (2023) berichtet über einen Versuch, ChatGPT für die Korpusanalyse emotionaler Sprache einzusetzen (dem Analysemodell von Schwarz-Friesel 2013 folgend). ChatGPT erkennt viele relevante sprachliche Elemente, weist aber insgesamt eine „mangelnde Bewertungsfähigkeit“ von Stilmitteln und insbesondere des Persuasionspotenzials von Texten auf (Burschik 2023, S. 11). Ist es nur eine Frage der Zeit, bis KI perfekte Emotionsana-

lysen durchführt und perfekte emotive Texte abliefer? Dazu will ich keine Prognose und auch keine Bewertung abgeben, aber die Tendenzen der Analyse in Kapitel 4 lassen vermuten, dass soziokulturelle Standards des Emotionserlebens durch generative KI weiter gefestigt werden.

Einige bereits etablierte Trends der linguistischen Forschung werden sich meiner Einschätzung nach durch den Einsatz von KI verstärken: die Entwicklung hin zu Digital Humanities und Big Data, zu Korpuslinguistik und quantitativen Methoden, automatisierten Auswertungen, komplexen Visualisierungen und die selbstverständliche und ebenfalls automatisierte Berücksichtigung von Multimodalität (gesprochene Sprache, interaktionale Videodaten, Bilder etc.). Als Geisteswissenschaftlerin plädiere ich dafür, anstelle der Abwertung der Geisteswissenschaften die Perspektive der Digital Humanities in die KI-beeinflusste Wissensgesellschaft einzubringen, insbesondere indem die Mitwirkung an der Lösung spezifischer Probleme verstärkt wird (vgl. Poschinger/Coon 2023). *AI Literacy* als mögliches neues Bildungsziel (Tseng/Warschauer 2023) könnte mit Critical Algorithm Studies (Sommerfeld 2022, S. 42-43) kombiniert werden, um den NutzerInnen Wissen über Fähigkeiten und Grenzen, Filterblasen und Undurchsichtigkeit von generativer KI sowie umfassende kritische Kompetenzen zu vermitteln. Die Komponente Emotion sollte dabei nicht vernachlässigt werden, da die sprachliche Ver- und Bearbeitung unseres emotionalen Erlebens wichtige gesellschaftliche Funktionen erfüllt, die weit darüber hinausgehen, die Stimmung von Konsumentinnen und Konsumenten besser einzuschätzen oder sie durch demonstrierte Emotionalität von KI-Anwendungen positiv zu beeinflussen.

## Literatur

- Asher, Nicholas & Zweigenbaum, Pierre (2020): Artificial Intelligence and language. In: Marquis, Pierre; Papini, Odile; Prade, Henri (Hg.): *A guided tour of Artificial Intelligence research. Volume 3: Interfaces and applications of Artificial Intelligence*. Cham: Springer, S. 117-145.
- Assunção, Gustavo; Patrão, Bruno; Castelo-Branco, Miguel & Menezes, Paulo (2022): An overview of emotion in artificial intelligence. *IEEE Transactions on Artificial Intelligence* 3 (6), S. 867-886.
- Auer, Peter (2013): Einleitung. In: Auer, Peter (Hg.): *Sprachwissenschaft: Grammatik – Interaktion – Kognition*. Stuttgart; Weimar: Metzler, S. 1-41.
- Benites, Fernando; Delorme Benites, Alice & Anson, Chris M. (2023): Automated text generation and summarization for academic writing. In: Kruse, Otto; Rapp, Christian; Anson, Chris M.; Benetos, Kalliopi; Cotos, Elena; Devitt, Ann & Shibani, Antoinette

- (Hg.): *Digital writing technologies in higher education. Theory, research, and practice.* Cham: Springer, S. 279-301.
- Bonfert, Michael; Spliethöver, Maximilian; Arzaroli, Roman; Lange, Marvin; Hanci, Martin; Porzel, Robert (2018): If you ask nicely. A digital assistant rebuking impolite voice commands. *ICMI '18: Proceedings of the 20th ACM International Conference on Multimodal Interaction*, October 2018, S. 95-102. DOI: <https://doi.org/10.1145/3242969.3242995>.
- Brinker, Klaus; Cölfen, Hermann; Pappert, Steffen (2014): *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden.* 8., neu bearbeitete und erweiterte Aufl. Berlin: Schmidt (Grundlagen der Germanistik 29).
- Bubeck, Sébastien; Chandrasekaran, Varun; Eldan, Ronen; Gehrke, Johannes; Horvitz, Eric; Kamar, Ece; Lee, Peter; Lee, Yin Tat; Li, Yuanzhi; Lundberg, Scott; Nori, Harsha; Palangi, Hamid; Tibeiro, Marco Tulio & Zhang, Yi (2023): Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv:2303.12712*. Abrufbar unter: <https://doi.org/10.48550/arXiv.2303.12712>.
- Burschik, Nele (2023): Sprache im Zeitalter Künstlicher Intelligenz. Ist ChatGPT die Schreibmaschine der Zukunft? (Dossier). *Sprachreport* 39 (3), S. 8-16.
- Butz, Andreas & Krüger, Antonio (2017): *Mensch-Maschine-Interaktion.* 2., erweiterte Aufl. Berlin, Boston: De Gruyter (De Gruyter Studium).
- Deppermann, Arnulf (2018): Sprache in der multimodalen Interaktion. In: Deppermann, Arnulf & Reineke, Silke (Hg.): *Sprache im kommunikativen, interaktiven und kulturellen Kontext.* Berlin, Boston: De Gruyter (Germanistische Sprachwissenschaft um 2020 3), S. 51-85.
- Devedzic, Vladan (2022): Identity of AI. *Discover Artificial Intelligence* 2 (23), 19 S. <https://doi.org/10.1007/s44163-022-00038-0>.
- Dong, Chenhe; Li, Yinghui; Gong, Haifan; Chen, Miaoxin; Li, Junxin; Shen, Ying & Yang, Min (2022): A survey of Natural Language Generation. *ACM Computing Surveys* 1 (1), S. 1-38. <https://arxiv.org/pdf/2112.11739.pdf> [Stand vom 22-09-2023].
- Gelitz, Christiane (2022): Interview mit einem Cog. In: Bischoff, Manon (Hg.): *Künstliche Intelligenz. Vom Schachspieler zur Superintelligenz?* Berlin: Springer, S. 93-103.
- Girgensohn, Katrin & Sennewald, Nadja (2012): *Schreiben lehren, schreiben lernen: Eine Einführung.* Darmstadt: WBG (Einführungen Germanistik).
- Guzman, Andrea L. & Lewis, Seth C. (2020): Artificial intelligence and communication: A Human-Machine Communication research agenda. *New Media & Society* 22 (1), S. 70-86. DOI: <https://doi.org/10.1177/1461444819858691>.
- Hasenbein, Melanie (2023): *Mensch und KI in Organisationen. Einfluss und Umsetzung Künstlicher Intelligenz in wirtschaftspsychologischen Anwendungsfeldern.* Berlin: Springer.

- Kopp, Stefan & Krämer, Nicole (2021): Revisting Human-Agent Communication: The importance of joint co-construction and understanding mental states. *Frontiers in Psychology* 12, Article 580955. DOI: 10.3389/fpsyg.2021.580955.
- Kövecses, Zoltán (2000): *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge: Cambridge University Press/Éditions de la Maison des Sciences de l'Homme (Studies in emotion and social interaction: Second series).
- Laar, Franziska (2023): æffekt: Emotionale Wirkungsforschung mithilfe von Künstlicher Intelligenz. In: Schweppenhäuser, Gerhard; Popp, Judith-Friederike & Bauer, Christian (Hg.): *Designmethoden im Zeitalter ihrer technischen Reproduzierbarkeit*. Wiesbaden: Springer VS, S. 63-97.
- Lee, Sangwon; Lee, Naeun & Sah, Young June (2020): Perceiving a mind in a chatbot: Effect of mind perception and social cues in co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction* 36 (10), S. 930-940.
- Lind, Miriam (2022): „Alexa, 3, Sprachassistentin, hat die Religion für sich entdeckt“. Die sprachliche Anthropomorphisierung von Assistenzsystemen. In: Lind, Miriam (Hg.): *Mensch – Tier – Maschine. Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*. Bielefeld: Transcript (Human-Animal Studies 24), S. 347-370.
- Lotze, Netaya (2022): Zur Adressierung des Unbelebten – Grenzen von pragmatischer Konzeption. In: Lind, Miriam (Hg.): *Mensch – Tier – Maschine. Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*. Bielefeld: Transcript (Human-Animal Studies 24), S. 305-325.
- Manzeschke, Arne & Assadi, Galia (2020): Emotionen in der Mensch-Maschine-Interaktion. In: Heßler, Martina; Liggieri, Kevin (Hg.): *Technikanthropologie. Handbuch für Wissenschaft und Studium*. Baden-Baden: Nomos/Edition Sigma, S. 165-171.
- Marichalar Quezada, Rebeca; Bartl, Michael & Garrecht, Gabriel (2021): Emotion AI. Neue Formen der Emotionsmessung durch Künstliche Intelligenz. In: Lichtenhaler, Ulrich (Hg.): *Künstliche Intelligenz erfolgreich umsetzen*. Wiesbaden: Springer Gabler, S. 205-223.
- Mavrina, Lina; Szczuksa, Jessica; Strathmann, Clara; Bohnenkamp, Lisa Michelle; Krämer, Nicole & Kopp, Stefan (2022): „Alexa, you're really stupid“: A longitudinal field study on communication breakdowns between family members and a voice assistant. *Frontiers in Computer Science* 4 (Article 791704). DOI: 10.3389/fcomp.2022.791704.
- Misselhorn, Catrin (2021): *Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern und Co.* Ditzingen: Reclam (Was bedeutet das alles?).
- Ortner, Heike (2014): *Text und Emotion. Theorie, Methode und Anwendungsbeispiele emotionslinguistischer Textanalyse*. Tübingen: Narr (Europäische Studien zur Textlinguistik 15).
- Parry, Anat (2023): AI will never convey the essence of human empathy. *Nature Human Behaviour* 7, S. 1808-1809. DOI: <https://doi.org/10.1038/s41562-023-01675-w>

- Petz, Gerald (2019): *Opinion Mining im Web 2.0. Ansätze, Methoden, Vorgehensmodell*. Wiesbaden: Springer Gabler.
- Pins, Dominik; Boden, Alexander & Essing, Britta (2020): „Miss Understandable“ – A study on how users appropriate voice assist and and deal with misunderstandings. *MuC'20*, September 6-9, 2020, Magdeburg, Germany. DOI: <https://doi.org/10.1145/3404983.3405511>.
- Philipp, Maik (2015): *Schreibkompetenz: Komponenten, Sozialisation und Förderung*. Tübingen: Francke (UTB 4457).
- Poschinger, Felix & Coon, Robert (2023): Toward a semiotic pyramid: language studies, AI, and knowledge exchange economy. *Language and Semiotic Studies* 9 (3), S. 393-407. DOI: <https://doi.org/10.1515/lass-2023-0016>.
- Rickheit, Gert; Weiss, Sabine & Eikmeyer, Hans-Jürgen (2010): *Kognitive Linguistik: Theorien, Modelle, Methoden*. Tübingen, Basel: Narr (UTB 3408).
- Robinson, Bradley (2023): Speculative propositions for digital writing under the new autonomous model of literacy. *Postdigital Science and Education* 5, S. 117-135. DOI: <https://doi.org/10.1007/s42438-022-00358-5>.
- Rupprecht, Werner (2014): *Einführung in die Theorie der kognitiven Kommunikation. Wie Sprache, Information, Energie, Internet, Gehirn und Geist zusammenhängen*. Wiesbaden: Springer Vieweg.
- Schwarz-Friesel, Monika (2013): *Sprache und Emotion*. 2., aktualisierte und erweiterte Aufl. Tübingen; Basel: Francke (UTB 2939).
- Sommerfeld, Alicia (2022): KI Made in Europe. Diskursive Unhintergehrbarkeiten und deren Potenziale für eine angewandte Rhetorik. *Rhetorik* 41 (1), S. 40-57. <https://doi.org/10.1515/rhet-2022-0006>.
- Thar, Evelyn (2015): „Ich habe Sie leider nicht verstanden.“ *Linguistische Optimierungsprinzipien für die mündliche Mensch-Maschine-Interaktion*. Bern [u.a.]: Lang (Sprache in Kommunikation und Medien 8).
- Tseng, Waverly & Warschauer, Mark (2023): AI-writing tools in education: if you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*. DOI: <https://doi.org/10.1515/jccall-2023-0008>.
- Udrea, Titus; Fuchs, Daniela & Preissl, Walter (2022): *Künstliche Intelligenz. Verstehbarkeit und Transparenz*. Endbericht einer Studie des Instituts für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften. Wien: Österreichische Akademie der Wissenschaften. Abrufbar unter URL: [https://epub.oeaw.ac.at/0xc1aa5576\\_0x003d5f4a.pdf](https://epub.oeaw.ac.at/0xc1aa5576_0x003d5f4a.pdf) [Stand vom 22-09-2023].
- Wyss, Stefanie (2023): KI-Schreiben. *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. Berlin, Boston: De Gruyter. Abrufbar unter: [https://www.degruyter.com/database/WSK/entry/wsk\\_22316506/html](https://www.degruyter.com/database/WSK/entry/wsk_22316506/html) [Stand vom 11-09-2023].

# **Zwischen AI Act und Posthumanismus. Künstliche Intelligenz und ihre ethischen „Radialkräfte“**

**Michael Funk**

## *Zusammenfassung*

Ethik feiert Konjunktur. Neuerdings dreht sich alles um Künstliche Intelligenz. Die alte Frage „Ethik – wofür braucht man das?“ ist auf dem absteigenden Ast. Längst hat der Run um die Deutungshoheit bei der Regulierung von KI begonnen. Ethik ist Richtlinie, politisches Programm, wirtschaftliches Interesse und/oder eine humanistische Gutenachtgeschichte, die es in der Morgendämmerung intelligenter Maschinen zu überwinden gilt. In vorliegendem Beitrag wird die Metapher der „Radialkraft“ verwendet, um vier ausgewählte Spannungsfelder aufzuzeigen. Was macht Ethik aus? Welche Wissensdiversität hält Ethik in Zeiten Künstlicher Intelligenz im Gleichgewicht? Diese Fragen diskutiere ich anhand von zwei Beispielen: dem EU AI Act sowie der posthumanistischen Kritik an menschenzentrierter Ethik.

## **Einleitung: Von Ethiken zur Wissensdiversität**

Räumen wir gleich zu Beginn zweierlei ein: Erstens, „Künstliche Intelligenz“ (KI) ist eine schiefen Übersetzung von „artificial intelligence“ und sollte eigentlich „technische Informationsverarbeitung“ meinen. Als solche umfasst sie eine Vielzahl verschiedener Produkte und Verfahren – etwa des *machine learning* als Mittel zum Zweck der Mustererkennung in Daten und anschließender Prognose realer Ereignisse. KI ist Querschnittstechnologie in vielen Lebenslagen. In vorliegendem Beitrag soll es nicht um die technischen Details gehen, sondern um die ethischen Radialkräfte durch deren Einsatz. Dabei werde ich der eingebürgerten Redeweise von KI folgen (vgl. auch die Beiträge in diesem Band).

Zweitens: So wie es nicht *die* KI gibt, so gibt es auch nicht *die* Ethik. Hier treffen sich zwei Metaphern. Zum einen KI als Bild für Technik, die wie Menschen oder Tiere „intelligent“ sein könnte. Zum anderen klingt der Titel des vorliegenden Beitrages doch arg nach Computersimulation von Maschinenbauteilen, also nach dem Bild der „Radialkräfte“ die im Verbund mit „Fliehkräften“ auf einen rotierenden Körper wirken. Was da rotiert, nennen wir Ethik. Vermutlich entspricht das Bild der Fliehkräfte eher der Erwartung: KI erzeugt Fragen und Probleme, welche die Ethik aus ihrem altehrwürdigen Zentrum katapultieren – z.B. moralische Maschinen oder künstliches Bewusstsein als Grundlage für Robotergesetze. Man kann das aber auch herumdrehen: Vielleicht sind es ja doch Radialkräfte, die Ethik in Zentrumnähe halten – z.B. durch Regulierung und Kodizes oder die Entmystifizierung technologischer Verheißungen (auch KI ist „nur“ technische Informationsverarbeitung). Dabei kann Ethik in dreifacher Hinsicht auf bekannten Unterscheidungen aufbauen<sup>1</sup>:

1. **Moral/Sitte** im Sinne gewohnter Handlungen, die sich stark zwischen menschlichen Gruppen unterscheiden können, einschließlich taubem Fanatismus oder der Sitte, die Möglichkeit von Moral überhaupt zu leugnen, bis hin zu Fragen nach den Grenzen und der Verbindlichkeit individueller oder gemeinschaftlicher sozialer Orientierungen (z.B. Hoerster 2022; Somek 2021; Williams 1986)

## 2. **Ethik**

a) als rationale **Moralwissenschaft**, Suche nach der Begründung des in sich Guten – im Gegensatz zum bloß zweckdienlich Nützlichen (z.B. Horster 2009; Murdoch 2023; Tugendhat 1984) – bzw. des spezifisch menschlich Guten und dessen Natur (z.B. Fischer 2022; Foot 2023; Kallhoff 2022), sowie über das Konzept einer modernen Wissenschaft hinaus:

b) als **angewandte Ethik und ethische Praxis** jenseits theoretischer Begründungsdebatten oder formaler Metaethik (Johnson/Toulmin 1988; Schwemmer 1980; Thurnherr 2010)

c) und noch stärker als **Lebenskunst/Klugheitslehre** des glücklichen individuellen und gemeinsamen Lebens, wie schon in der Antike oder außereuropäischen Traditionen (Hossenfelder (Hrsg.) 2013; Masek 2023; Nussbaum 2020)

3. **Kodizes/Ethikrichtlinien**, in denen Handlungsnormen explizit ausformuliert sind – wie im hippokratischen Eid oder den Zehn Geboten – im Unter-

---

1 Zur vertieften Analyse mit Blick auf die Roboter- und KI-Ethik siehe Funk (2022a).

schied zu moralischem Handlungswissen, das auch implizit durch Bräuche und Rituale ausgedrückt wird.

Die Frage des vorliegenden Beitrags lässt sich entsprechend zusammenfassen:

*Wie verändert der Umgang mit „technischer Informationsverarbeitung“ die Ethik im dreifachen Sinne: hinsichtlich moralischer Gewohnheiten, moralwissenschaftlicher Reflexion (einschließlich Lebenskunst sowie ethischer Praxis) und Ethikrichtlinien?*

Ich möchte diese Frage zuerst mit einer Übersicht zu vier „Radialkräften“ bearbeiten und anschließend anhand von zwei Beispielen vertiefen: dem AI Act sowie der kritischen Diskussion über die Menschenzentriertheit der Ethik. Dabei schließe ich an meinen Beitrag *Nichtwissen ist Macht! Künstliche Intelligenz und ihre diversen Ethiken* (vgl. Funk 2024) unmittelbar an, wo zwei Formen der Wissensdiversität unterschieden wurden:

**Wissensdiversität I:** Vielfalt moralischer Lebensstile (oben 1.) und ethischer Lebenskünste (2.c)

- bildet den Fundus personalen Wissens und zwischenmenschlicher Fertigkeiten: erlernt durch Imitation, Tradition, zwischenmenschliche Interaktionen, Trial and Error, pragmatische Bewährung durch wiederholt gelingenden Umgang, Riten, Sitten, Bräuche, Intuition, Moral Sense, Umgehen mit Emotionen, leiblichen Impulsen etc.
- Kritische Infrastruktur freiheitlich-demokratischer Gesellschaften
- wird auch in der Moralphilosophie, Metaethik und Moralbegründung stillschweigend vorausgesetzt
- Prinzipielle Grenze der KI: Maschinen können das (höchstens teilweise) simulieren, aber nicht authentisch leben

**Wissensdiversität II:** Vielfalt der Wissensformen, die in der Ethik (2.) wirksam werden

- s.o. Wissensdiversität I (denn Ethik wird von lebenden Menschen gemacht)
- Wissensprozesse, epistemische Praxen: Verfahrenswissen der angewandten Ethik (Kasuistik, Einzelfallanalysen, Case Studies etc.), Problemlösungsstrategien ethischer Praxis (logisch-ableitend, analogisch-vergleichend, top-down, bottom-up etc.) (2.b)

- propositionales, formales Wissen der Metaethik: Sprachanalyse ethischer Argumente und moralischer Urteile, formale Logik ethischer Begründungen (2.a)
- Kenntnis der Positionen der Moralbegründung: deontologische Ethik, Utilitarismus, Diskursethik, Tugendethik etc. (2.a)
- propositionales Wissen ethischer Theorien: Konzepte, Werke, AutorInnen auswendig lernen (2.a)
- Bsp.: Umgang mit Fehlschlüssen (z.B. Sein-Sollen-Fehlschluss oder humanitaristischer Fehlschluss)
  - auswendig kennen (propositionales Wissen)
  - moralphilosophisch korrekt in Texten oder Argumenten erkennen, aufdecken und korrigieren können
  - im moralischen Alltag wahrnehmen und vermeiden (oder bei nicht erheblicher Problematik gelassen hinnehmen) können

Die These, an die in vorliegendem Beitrag angeschlossen wird, lautet:

*Ethik Künstlicher Intelligenz ist für leiblich-sinnliche Menschen, für emotionale Wesen im kommunikativen Gemeinschaftsleben, nicht jedoch für mathematisch reduzierte Massenpunkte oder Vektoren in Computermodellen.*

Diesen Unterschied zu erkennen, setzt eine Haltung voraus, die in den aktuellen Debatten um KI-Ethik nicht selbstverständlich ist.<sup>2</sup> Eine wesentliche Ursache ist die Verennahmung kritischer Ethik durch industrielle Interessen im so bezeichneten Überwachungskapitalismus. Insofern finden wir hier eine der stärksten Radialkräfte auf die aktuelle Ethik. Zu den Vorurteilen im Umgang mit KI gehört, die Diversität ethischen und moralischen Wissens zu unterlaufen. Das passiert z.B., wenn Ethik auf Guidelines reduziert wird, also „Checklistenökonomie“, wonach moralische Verantwortung im Moment des Abhakens von Kästchen zur Fairness und Vertrauenswürdigkeit erledigt ist (erstes Beispiel). Der „Export“ europäischer Ethikrichtlinien ist nicht wertfrei und könnte ungewollt zur Wiederholung kolonialer Handlungsmuster führen. Dazu gehört das Ausblenden moralischen Handlungswissens indigener Kulturen zugunsten bestimmter Menschen- und Weltbilder. Vielleicht ist die Menschenzentriertheit bestimmter anthropozentrischer Ethiken ein Gift für holistische Alternativen (zweites Beispiel)? In einer Art Prisma sollen **kulturell-soziale Radialkräfte der Ethik** aufgezeigt werden, so wie sie vom Umgang mit KI ausgehen. Dabei muss zur oben genannten Forschungs-

2 Siehe zur kritischen Diskussion z.B.: Coeckelbergh 2022; Dignum 2019 oder Eubanks 2018. Ich schließe auch an meine Analyse in Funk 2023 an.

frage noch ergänzt werden, dass sie natürlich ein Framing vorgibt. Sie ließe sich auch herumdrehen: Welchen Einfluss hat der Umgang mit Ethik (in ihren drei Bedeutungen) auf KI? Beide Perspektiven gehören zwar zusammen. Jedoch gebe ich der Frage nach dem Einfluss von KI auf Ethik hier den Vorzug, da sie im Anbetracht wirtschaftlich-technologischer Tatsachen und hinterherhinkender Debatte über das, was wir wollen, die drängendere zu sein scheint.

## Radialkräfte der Ethik

Sehen wir uns einige der Radialkräfte an. Bildlich schauen wir auf die Kufen einer Eiskunstläuferin und die wirkende Energie. Welche Kraft hält sie in der Spur, wenn sie sich virtuos in die Kurven legt? Ganz ähnlich ist es – metaphorisch – mit der Ethik. Und alle hoffen, dass sie nicht stolpert und sich erst recht nicht dabei verletzt. Ich möchte vier dieser Spannungsfelder herausheben – sicherlich lassen sich noch mehr finden. Sie beschreiben eine allgemeine Verfasstheit und werden hier mit Blick auf KI wiedergegeben.

I. Unter dem Titel „**Nehmt der Industrie die Ethik weg!**“ (Metzinger 2019) lässt sich eine erste Radialkraft zusammenfassen. Das Problem liegt darin, dass Ethik durch den industriellen Umgang mit KI völlig aus dem Gleichgewicht gerissen wird. Damit sind weniger neue Fragen im Anbetracht neuer technologischer Möglichkeiten gemeint. Denn diese geben ja der ethischen Diskussion Schwung. Auch sollten Beiträge zur ethischen Bewertung aus der technischen, industriellen Praxis heraus ausdrücklich erwünscht sein! Es wäre doch vermessen, Ethik zum elitären akademischen Sport zu machen, jenseits des Alltagsverständes. Problematisch ist jedoch die Umnutzung von Ethik zur Interessenvertretung. Es wirken also zwei Pole aufeinander ein:

1. Fliehkraft: industrielle oder politische Lobbyarbeit, „ethics washing“ zur Akzeptanzbeschaffung neuer KI, Ausspielen von Ethik gegen politische Regulierung, Unternehmen mit marktbeherrschender Stellung nutzen Regulierung (juristische und/oder ethische), um MitbewerberInnen klein zu halten
  - ⇒ Umnutzen von Ethik, Ethik als zweckdienliches Mittel zur Interessenvertretung
2. Gegenpol: ethischer Elfenbeinturm
  - ⇒ **Radialkraft:** Balanceakt der KI-Ethik zwischen bodenständiger Praxisrelevanz sowie autonomer Suche nach in sich Gutem und Schutz vor Missbrauch

- ⇒ Angewandte Ethik (2.b): Vermittlung ethischen Fachwissens und alltäglichen Handlungswissens

Eine Lösungsstrategie besteht in wirksamer Wissensdiversität durch authentische transdisziplinäre Integration: industrielles, technisches Wissen im Dialog mit politischem Wissen und wissenschaftlichem, ethischem Fachwissen. Die Diversität ethischen Wissens, von gelebter Moral (1.) bis hin zur rationalen Argumentation (2.) sollte nicht durch Lobbyismus entwertet werden.

II. Die zweite Radialkraft ist verbunden mit „**Checklistenethik**“ (Nowotny 2023, S. 214). Dabei geht es weniger um den Missbrauch ethischer Etiketten zur Akzeptanzbeschaffung, sondern mehr um die Reduktion ethischen Wissens auf Kodizes (3.):

1. Fliehkraft: Reduktion von Ethik auf KI-Guidelines, die top-down vorgegeben werden und dann „nur“ noch abgehakt werden müssen (Regulierung von oben)
  - ⇒ Reduktion von Ethik auf eine spezifische Wissensform
2. Gegenpol: bottom-up Reduktion von Ethik auf das je individuelle Befinden
  - ⇒ **Radialkraft:** Balanceakt der KI-Ethik zwischen gemeinschaftlich verbindlichen Guidelines (3.) und konkreter Lebensführung (1., 2.c): auch Guidelines wollen von Menschen gelebt sein
  - ⇒ Angewandte Ethik: Vermittlung von universellen Richtlinien und spezifischen Alltagsfällen (Case Studies, Kasuistik)

Auch hier ergeben sich Lösungsstrategien durch transdisziplinäre Integration. Vielleicht ist das offensichtlicher, wenn von technischen oder geisteswissenschaftlichen Fächern die Rede ist. Jedoch gibt es auch subtilere Formen der Integration, konkret von Wissensdiversität II (s.o.): propositionales Wissen (explizite Formulierungen von Guidelines) in Verbindung mit moralischem Alltagswissen, ethisch-reflexivem Fachwissen sowie Ethik als Lebenskunst. Transdisziplinär ist eine solche Integration weniger wegen der Berufsabschlüsse der beteiligten Menschen, sondern wegen der verschiedenen Erfahrungshorizonte.

III. Bleibt noch der „**Elfenbeinturm**“ philosophischen, weltabgewandten ExpertInnenstums, der eine eigene Hybris ausprägt. Da ist zum einen die weltabgewandte Träumerei oder Unverständlichkeit verkopfter Glasperlenspiele, die zuweilen Hand in Hand geht mit Angst vor dem wahren, sinnlich-körperlichen Leben. So folgt entweder die Furcht vor überlegenen Maschinen oder die schlichte Belanglosigkeit, wenn es um das Lösen

realer Herausforderungen geht. Zum anderen wirkt die philosophische Selbstbeschränkung fatal, wenn sie im formal-geistigen Kunstraum nur noch das als Ethik übrig lässt, was durch KI besonders gut ersetzt werden kann – z.B. metaethische Theorien:

1. Fliehkraft: ethisches ExpertInnentum in einer eigenen Bubble über den Köpfen anderer gesellschaftlicher Gruppen und/oder skeptisches Alleszer-zweifeln
  - ⇒ Angst vor KI und dem Gespenst überlegener Maschinen
2. Gegenpol: Entwissenschaftlichung von Ethik, unreflektiertes Aufeinander-prallen moralischer Vorstellungen und/oder blinder, sturer Dogmatismus bis hin zur KI-Ideologie, Transhumanismus als Mythologie und Weltbild, Glaube an Superintelligenz
  - ⇒ **Radialkraft:** selbstkritische Reflexion auf die Chancen und Grenzen wissenschaftlichen Wissens und Ethik als fluider, lebenslanger, zwischen-menschlicher Lernweg (Wissen als Klugheit (2.c)) gegen den Kokon des Formalismus oder dogmatische Elitendünkel
  - ⇒ Angewandte Ethik: skeptisches Prüfen und authentische Haltung im Um-gang mit KI verbinden

Transdisziplinäre Integration kann sich dann sogar als überlebenswichtig für die akademische, philosophische Ethik herausstellen. Schafft sie es, sich konstruktiv in den Debatten zum Umgang mit KI einzubringen, oder reduziert sie sich selbst auf einen ethischen Automaten, der früher oder später von einer entsprechend leistungsfähigen Maschine – völlig zu Recht – ersetzt wird?

IV. Nicht weit weg vom Elfenbeinturm ist das vierte Spannungsfeld der „**narzisstischen Bettgeschichten**“ (Barad 2020, S. 97). Es geht also um die kritische Auseinandersetzung mit den Menschenbildern des Humanismus – die vielleicht viel zu europäisch und männlich geprägt sind – zugunsten alternativer Sichtweisen. Posthumanismus meint nicht die Überwindung der menschlichen Natur mit technologischen Mitteln (das wäre Transhumanismus), sondern die Überwindung des Humanismus:

1. Fliehkraft: dogmatische weltanschauliche Anthropozentrik und ein sich darin erschöpfernder Humanismus, weiterhin Eurozentrik und Androzen-trik
  - ⇒ „narzisstische“ Wissensform, „blind“ für Alteritäten
2. Gegenpol: Elimination jeder Anthropozentrik, die Menschenzentriertheit der Ethik beenden bzw. das europäische und/oder männliche Menschen-bild überwinden

- ⇒ **Radialkraft:** skeptische Aufklärung und kritischer Dialog
- ⇒ Angewandte Ethik: Fingerspitzengefühl, wirkliche Diskriminierung zu erkennen, sich nicht von Strohmannargumenten ablenken lassen, trotzdem Menschen, ihre Endlichkeit und Verletzlichkeit im Zentrum der Ethik zu erkennen

Im Sinne einer transdisziplinären Integration bietet sich hier der skeptische Dialog als methodische Arbeit an. Es passiert ja in der Tat viel zu schnell, dass unabsichtlich zu enge Menschenbilder zu Schäden führen. Ziel könnte moralische Klugheit sein, also authentisches, sinnlich-leibliches Wissen im Umgang mit Alterität sowie selbstreflexives Wissen im Umgang mit der eigenen Sprache. Es ließen sich weitere Radialkräfte ergänzen, die Problemfelder sind komplex. Doch sehen wir uns im Fortgang einfach zwei Beispiele an.

### **Erstes Beispiel: AI Act und Verantwortung als Richtlinie**

Im Mittelpunkt des EU AI Act steht die juristische Regulierung von **inakzeptablen Hochrisiko-KI-Systemen** unter der **Maßgabe vertrauenswürdiger und fairer Technologieentwicklung**. Mit den Grundrechten der Bevölkerung soll gleichzeitig die EU als Wirtschaftsstandort gestärkt werden.

„The aim of the new rules is to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models.“  
(European Parliament 2024)

HerstellerInnen dürfen nach einheitlichen Maßstäben nur Produkte unter Achtung europäischer Prinzipien und Werte auf den Markt bringen (Schmid et al. 2023, S. 237-238). Auf Grundlage Harmonisierter Europäischer Normen (HEN) durchlaufen Produkte ein „Konformitätsbewertungsverfahren“, sodass auch für HerstellerInnen auf dem europäischen Binnenmarkt Rechtssicherheit gegeben ist. Regulatorische Hürden sollen innovative Technologiegestaltung nicht behindern. In den Gesetzestexten sind folglich auch Unterstützungsmaßnahmen enthalten. Ziel ist „eine gute Balance zwischen Risiken und Chancen“ sowie „ein Werkzeug einer im positiven Sinn nachhaltigen Entwicklung“ (ebd., S. 252). Grundlage ist die **Transparenz und Erklärbarkeit** funktionaler Eigenschaften von KI-Systemen, um Vertrauenswürdigkeit auf NutzerInnenseite zu ermöglichen. Drei Spezifika der Technikfolgenabschätzung werden übernommen (Grunwald 2003; Grunwald 2017):

1. Die Bewertung erfolgt **anwendungsspezifisch**. Zur Analyse der konkreten Anwendung werden klassifikatorische Merkmale vorgeschlagen, die sich an der Ersetzung menschlicher Handlungen durch konkrete KI-Methoden orientieren (Schmid et al. 2023, S. 93-102, S. 109-121).

2. Der eingeschränkte Satz funktionaler Optionen steht von vorneherein einer Operationalisierung von Ethik in einer KI entgegen. Im AI Act nimmt entsprechend die **menschliche Aufsicht bzw. Überwachung** der Technologien eine zentrale Rolle ein. Im Mittelpunkt steht der Schutz von Grundrechten (ebd., S. 238-239). Dahinter steckt die in der Technikphilosophie bekannte Einsicht:

„Im moralischen Sinne wird das computerisierte Entscheidungssystem, auch wenn es hinreichend lernfähig ist, dadurch nicht zu einer vollverantwortlichen moralischen Person als Träger der moralisch-praktischen Vernunft gemacht werden können.“ (Lenk 1994, S. 82-83)

3. **Risikomanagement** und **Risikoabschätzung** dienen der Erhöhung der Produktsicherheit unter Einbezug von **Zweck-Mittel-Analysen** (Anwendungsziele, Umnutzungs- bzw. Missbrauchspotenziale) (Grunwald 2017, S. 339-350; Grunwald (Hrsg.) (1999), S. 155-174; Lenk 2007 S. 166-197). Die Regulierung durch den AI Act zielt besonders auf Technologien mit „inakzeptablen“ Schadensrisiken für einzelne Menschen, Gemeinschaften oder Sachwerte ab. Geringe oder minimale Risiken soll mit allgemeinen Transparenz- anforderungen und Verhaltenskodizes begegnet werden (Schmid et al. 2023, S. 128-136, S. 239-242).<sup>3</sup>

Wie schon bei den Ethikrichtlinien von 2019 (European Union 2019) bewirkt der AI Act eine stärkere Wahrnehmung ethischer Themen. Ethik lässt sich leicht rechtfertigen mit Verweis auf die Vorgaben der EU. Das alte Argument – „Ethik, wofür braucht man das?“ – scheint obsolet. Nach der UNESCO 2021 (Unesco 2022) hat auch die UN im März 2024 eine Resolution zu sicherer, vertrauenswürdiger und nachhaltiger KI – für alle Menschen während des kompletten Lifecycles – verabschiedet (United Nations 2024). Dabei handelt es sich um top-down verfasste Kompromisspapiere, die Empfehlungen aussprechen und für Kritik sorgen:

---

3 Je nach Darstellung werden 3 bis 5 Risikoklassen unterschieden. Entscheidend sind die Kriterien, anhand derer die Einstufung vorgenommen wird. Siehe auch: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [Stand vom 28-07-2024].

„Selbstredend haben die UN-Diplomaten darauf geachtet, ihren Regierungen keine Steine in den Weg zum Schlachtfeld zu legen: Die KI-Resolution gilt ausdrücklich nicht für KI im Militärbereich. Immerhin steht auf der Tagesordnung der nächsten Generalversammlung eine UN-Resolution zu ‚Tödlichen Autonomen Waffensystemen‘. Ein Konsensbeschluss dazu ist unwahrscheinlich. Schon im vorbereitenden Ausschuss haben unter anderem Russland und Indien dagegengestimmt.“ (Sokolov 2024)

Mit Blick auf den AI Act wird kritisiert, dass zwar Risiken im Mittelpunkt stehen, jedoch die Reflexion auf **wünschenswerte Zukünfte** zu kurz kommt:

„The legislation’s strong focus on risk also neglects the potential benefits of a technology. Regulators should aim to increase the overall value that the use of a technology creates for the public, so both risks and benefits should matter.“ (Prainsack/Forgò 2024)

Zu den generellen Bedenken bei Top-down-Richtlinien – über den Köpfen der Betroffenen – gehört die **globale Vereinnahmung durch industrielle Interessen**:

„Self-regulating ethical frameworks allow companies to choose how to deploy technologies and, by extension, to decide what ethical AI means for the rest of the world. [...] To understand what is at stake, we must focus less on ethics and more on power.“ (Crawford 2019, S. 224, siehe auch Gerdés 2022)

Die Problemlage reicht bis hin zu wirtschaftskritischen, postkolonialen Themen (Mbembe 2021) und dem Anliegen, koloniale Missstände sowie Imperialismus nicht beim „**Export**“ von **Ethikrichtlinien** zu wiederholen. So wird schon seit Längerem die Bedeutung eines interkulturellen Kanons, feministischer Forschung und „minority studies“ als Teil technikphilosophischer Grundlagen behandelt (Ihde 1993, S. 148-151). Hinsehen und Zuhören lässt sich von der Umwelt- und Klimaethik lernen, denn hier sind die interkulturellen Herausforderungen einer Ethik mit globalem Problemaufriss längst bekannt (Birnbacher 2022; Roser/Seidel 2015): **Guidelines hören nicht zu**. Umso wichtiger sind die verschiedenen globalen Stimmen bei der Regulierung von KI (Eke et al. 2023).

Es lassen sich Parallelen zur Diskussion technischer Katastrophen und IngenieurInnenverantwortung der 1980er und 1990er Jahre erkennen (Bhopal, Tschernobyl, Sandoz, Challenger-Unglück in Lenk 1994; sowie Lenk 1982). Zur ethischen Wissensdiversität in den Radialkräften des Umgangs mit KI gehören Guidelines (3.) aber auch ganz

positiv als integraler Teil der IngenieurInnenverantwortung – jedoch im Verbund mit moralischem Können (1.):

„Moralische Urteilskraft und ehrliches Sich-Orientieren an den Folgerungen des Standesethos im Sinne bestmöglicher Aufgaben- und Pflichterfüllung ist bei allen diesen Abwägungsproblemen und besonders bei Uneinigkeiten über Wertprioritäten sowie bei Interessenkonflikten unerlässlich.“ (Lenk 1994, S. 106)

„Und weil nur allgemeinmenschliche Interessen zählen, lassen sich die Erfahrungen anderer Kulturen integrieren: Die über Gedanken der Menschenrechte vermittelte Vorstellung von Identität ist im Prinzip flexibel und lernfähig.“ (Höffe 2014, S. 135)

Wer KI-Ethik auf das Abhaken von Guidelines reduziert (Nowotny 2023, S. 214), verstößt gegen diese Flexibilität und schlussendlich gegen Menschenrechte.

### **Zweites Beispiel: „Narzisstische Bettgeschichten“**

Radialkraft IV. lässt sich anhand der Diskussion um **Anthropozentrik und Intelligenz** vertiefen. Es geht dabei um die Frage, inwiefern Menschen im Zentrum der Ethik bleiben sollten und ob in weiterer Folge KI ethische *agency* hat – also selbst ein ethischer Akteur ist. Im ersten Schritt steht die Diskussion um Vielfalt menschlicher Intelligenzformen – im Gegensatz zum eingeengten Spektrum von Intelligenztests:

„In Zukunft werden zahlreiche Disziplinen an der Erarbeitung einer Definition der Intelligenz beteiligt sein und noch weit mehr Interessengruppen an ihrer Vermessung und ihrem Einsatz.“ (Gardner 2002, S. 37)

Es ist ja eine *der* Fragen unserer Zeit, wenn wir „intelligence“ als Informationsverarbeitung ansprechen und gleichzeitig die Doppeldeutigkeit intelligenten Verhaltens in einem Computer zulassen. So wird im zweiten Schritt die Verengung von KI auf „Unternehmens-Intelligenz“ kritisiert. Viele andere Intelligenzformen, die gleichfalls für KI Pate stehen könnten, werden ignoriert (Bridle 2023, S. 21). Hierzu zählen Verflechtungen mit und Beziehungen zu anderen natürlichen Intelligenzen (Pilze, Wälder, Ökosysteme etc.) (ebd., S. 23) bis hin zu *agency*, Handlungsmacht oder „Person-Sein“ der zukünftigen Technologien (Ebd., S. 29).

„Das abstrakte Denken biologischer Gehirne war die Stütze für die Entstehung aller Kultur und Wissenschaft. Aber diese Tätigkeit [...] wird ein kurzer Vorläufer der leistungsfähigeren Erkenntnisvermögen des anorganischen, posthumanen Zeitalters sein.“ (Rees 2017, S. 38)

Es geht also im **Posthumanismus** nicht mehr um die Öffnung des Blicks für bisher weniger beachtete menschliche Intelligenzformen wie bei Gardner, sondern über den Menschen hinaus. James Bridle geht so weit, nicht nur in den Naturwissenschaften, sondern auch in der Ethik die Menschenzentriertheit als Problem darzustellen (Bridle 2023, S. 338-341).

Posthumanistisches Denken bildet als methodisch-aufklärerisches Motiv eine konstruktive Stimme in der Ethik. Es stellt skeptische Kritik zu Verfügung. Gleichzeitig birgt es die Gefahr der zwischenmenschlichen Diskriminierung oder des Dogmatismus. Sehen wir uns hierzu das viel beachtete Konzept des **agentiellen Realismus** von Karen Barad genauer an. Es gehört zu den Entwürfen neueren materialistischen Denkens (**New Materialism**) (Hoppe/Lemke 2021, S. 59-61). Dabei folgt Barad der Kritik an dualistischen Denkweisen, in denen Natur/Kultur, Mensch/Nichtmensch oder Frau/Mann als Gegensätze gesehen werden (Braidotti 2014; Haraway 2018; Tsing 2022). Es geht ihr um Performativität als

„Infragestellung der unüberprüften Gewohnheiten des Geistes, die der Sprache und anderen Repräsentationsformen mehr Macht bei der Bestimmung unserer Ontologien zugestehen, als sie verdienen“ (Barad 2020, S. 10).

Neue geistige Gewohnheiten sollten Dynamiken der materiellen Welt stärker einschließen. Die Autorin verbindet damit ein posthumanistisches Programm, das sich radikal von Humanismus und Anthropozentrik absetzen will.

Der methodische Gehalt ihres Anliegens lässt sich anhand folgender Passagen nachvollziehen: „Humanistische Umlaufbahnen“ sollen verlassen werden, womit eben der „Griff anthropozentrischer Kräfte“ in „zeitgenössischen Denkmustern“ gemeint ist. „Hochgeschätzte Ionisierungen“ gingen von **poststrukturalistischen AutorInnen** wie Michel Foucault oder Judith Butler aus (siehe auch zu Deleuze und Guatteri: Hubatschke 2024). Jedoch blieben diese gehemmt von „anthropozentrischen Überbleibseln“. Das wiederum soll in Barads „posthumanistischen, performativen“ Neuaustrichtungen überwunden werden. Der „Ausnahmezustand des Menschen“ ist „aufs Korn zu nehmen“ (Barad 2020, S. 10-13). In einer „agentiell-realistischen Ontologie“ wird Materie als dynamische Agenten (*agencies*) betrachtet. Sie prägt aktiv die Relationen der Welt so-

wie „intraaktive, diskursive“ Praktiken. Materie „handelt“ und hat „Geschichte“ (ebd., S. 14-22). Der methodische Gehalt liegt in einer **Kritik des isolierten, archimedischen BeobachterInnenstandpunktes**. Barad setzt sich hierzu in einer Fallstudie mit Niels Bohr und den erkenntnistheoretischen Folgen der Quantenphysik auseinander.

Jedoch kann die Autorin zum einen ihren radikalen Wunsch, ohne jede Anthropozentrik auszukommen, nicht durchhalten. Denn sie ist ja eine menschliche Wissenschaftlerin, die in menschlicher Sprache für andere Menschen schreibt, mit dem Ziel, deren Denkgewohnheiten zu verändern. Sie bleibt einer methodisch-sprachkritischen Anthropozentrik verhaftet. Mit Selbstironie kommentiert Barad ihr Vorhaben als Fußnote:

„Es wäre verwunderlich, wenn mein eigener Versuch, einen erfolgreichen ionisierenden ‚Quantensprung‘ aus der humanistischen-repräsentationalistischen Umlaufbahn heraus zu machen, nicht derselben Anziehungskraft zum Opfer fiele [...].“ (ebd., S. 103)

Sie deutet damit die methodische Reichweite ihres reflexiven Versuchs an, zumindest die Rolle menschlicher BeobachterInnen in den Naturwissenschaften kritisch einzuordnen.

Auf der anderen Seite öffnet Barad missverständliche Türen für ein dogmatisches Umkippen ihrer Ideen. So benutzt sie eine **technomorphe Sprache** („Quantensprünge“, „ionisieren“), wodurch Menschen als bloße Naturereignisse angesprochen sind – und das, ohne ihre Metaphern methodisch-kritisch einzuordnen. „Handlung“ und „Geschichte“ fallen aus dem Bereich menschlicher Verantwortung, indem sie zu Merkmalen allgemeiner materieller Prozesse werden. Das ist weder mit der methodischen Trennung von Sein und Sollen vereinbar, noch mit den Grundlagen moderner Rechtsordnungen (Beweisführung auf der Tatsachenebene und individuelle Schuldfrage laut geltender Gesetze). Hinzu tritt die Reduktion von Humanismus auf anthropozentrische Prämissen der modernen Physik. Es wird sogar so pauschal verfahren, dass in einem Handstreich gleichzeitig die normativ-liberalen Kräfte des Humanismus sowie das verbundene Bildungsanliegen zur Ermächtigung politischer Diskurse ausgeblendet werden (ebd., S. 16, S. 29, S. 70). Selbst wenn für Naturwissenschaften die Kritik treffen sollte, bliebe deren Gültigkeit für Gesellschaftswissenschaften eigens zu prüfen. Darüber hinaus schließt die Autorin, dass „wir“ im Vollzug in der Welt „für Markierungen auf Körpern verantwortlich“ sind (ebd., S. 88). Dieser ethische Gemeinplatz kippt in sein Gegenteil an der Stelle, wo Barad das „wir“ ausdrücklich auf nichtmenschliche materielle Agenten ausdehnt: „Leider schloss Bohr jedoch wichtige posthumanistische Implikationen dieser Einsicht mit seinem letztlich humanistischen Verständnis des ‚wir‘ kurz“ – so ihre Kritik am Ende der Fallstudie (ebd., S. 99).

Selbst wenn dieses pantheistische (?), jedenfalls materialistisch-physikalistische Plädyer für Verantwortung gut gemeint sein mag, trägt es doch eine gefährliche Last in sich: Von Menschen hergestellte Maschinen wären dann mindestens genauso wichtig wie andere Menschen.

*Folglich führen Barads neue „geistige Gewohnheiten“ dazu, Verantwortung an der Elektrik von Computern zu orientieren und reale Ungerechtigkeiten gegenüber anderen Menschen zu übersehen. Die auch als Feministin bekannte Autorin öffnet damit Denkweisen Tür und Tor, die doch eigentlich nicht in ihrem Interesse liegen dürften.*

Verantwortung wird zu Beliebigkeit und die Rolle der Politik in ihrem Szientismus zu schwach reflektiert (Hoppe/Lemke 2021, S. 74-79). Verschärft wird diese Hybris durch die vorsorgliche Abfertigung alternativer Konzepte:

„Die einzigen Möglichkeiten [zur Konzeptualisierung der Materie] scheinen die Naivität des Empirismus oder dieselben alten **narzisstischen Bettgeschichten** zu sein [außer man vertrete eine] posthumanistische Sichtweise von Performativität [...].“ (Barad 2020, S. 97, Hervorhebung v. M.F.)

An Punkten wie diesen setzt sich Barad ähnlicher Kritik aus wie AutorInnen, die sie in den Schubladen des Humanismus, der Aufklärung oder Anthropozentrik pauschal anprangert: Stigmatisierung durch fehlendes Hinsehen und Zuhören, schlussendlich Ausblenden der verwickelten Vielfältigkeiten humanistischer Lebensspraxen.

Auch Rosi Braidotti gilt als prägende Vertreterin des **feministischen, kritischen Post-humanismus** (Hoppe/Lemke 2021, S. 83). Ihr geht es im Besonderen um den Widerspruch gegen Eurozentrismus, Imperialismus, Kolonialismus und die Annahme eines einheitlichen Subjekts. Sie formuliert einen monistischen Ansatz, demzufolge ein „Paradigmenwechsel“ in „Begriffen, Methoden und politischen Praktiken“ der „Gesellschaftstheorien“ erfolgen muss. Im Anschluss an Bruno Latour soll ihre „Agenda der posthumanen Verwicklung“ einen methodischen Anspruch einlösen und als Instrumentarium zum Verständnis menschlicher wie nichtmenschlicher Akteure im lebendigen Ökosystem der Erde dienen (Braidotti 2014, S. 6-11). Zentral ist hierfür die Verabschiebung des Anthropozentrismus aus den Geisteswissenschaften (ebd., S. 61-72, S. 147-149). Die binäre Logik aus Selbst und Anderem soll überwunden werden. Im antihumanistischen Programm ist kein Platz für „den Menschen“ (ebd., S. 19-31, S. 172-176), dementsprechend ist Posthumanismus darüber hinaus mit der Suche nach alternativen Formen des menschlichen Subjekts befasst (ebd., S. 42-50).

Es geht um die politische Vision der „Wiederherstellung eines panhumanen kosmopolitischen Bands“, das in Folge eines „selbstherrlichen“, „eurozentrischen“ Humanismus verloren gegangen sei (ebd. S. 57-59). Hierzu ist eine „radikale Entfremdung von den Begriffen moralischer Rationalität, einheitlicher Identität, transzendentalen Bewusstseins oder angeborener und allgemeiner moralischer Werte“ nötig. „Postanthropozentrischer Posthumanismus“ meint eine „Ökosophie“, „die die vielfältigen Schichten des Subjekts von innen nach außen über alles, was dazwischen liegt, transversal zu durchziehen sucht“ (ebd. S. 97). Als kritische Sicht auf die Gefahren selbstherrlicher Menschenbilder eröffnet Braidottis Entwurf wichtige Perspektiven und kann als Mahnung zur Erarbeitung einer nicht eurozentrischen KI-Ethik verstanden werden. Jedoch besteht auch hier die Gefahr, dass durch die Inklusion von technischen Mitteln wie Computern in einen ökosophischen Monismus eben gerade nicht tolerantes und offenes politisches Verhalten folgt:

„Die so verstandene ‚Maschinenwerdung‘ bezeichnet und realisiert die relationalen Kräfte eines Subjekts, das nicht mehr in einen dualistischen Rahmen gezwängt ist, sondern in einer bestimmten Verbindung zu vielfältigen Anderen steht und in seiner technologisch vermittelten planetarischen Umwelt aufgeht.“ (ebd., S. 96)

Die Verwicklung mit Maschinen kann dann – wenn Menschen sich nicht als ihre HerstellerInnen herausheben dürfen – zum fatalen Kurzschluss führen:

*Es wäre niemand mehr für die Technologien verantwortlich, Ungerechtigkeit zwischen Menschen wäre kein normatives Übel mehr, sondern eine bloße relationale Kraft, der nicht mehr mit rationaler Kritik und zwischenmenschlicher Courage begegnet werden könnte.*

Die destruktiven Potentiale materieller, technologischer Prozesse kann die Autorin durch ihren pauschalen Wurf nicht einfangen (Hoppe/Lemke 2021, S. 96-97). So erfrischend methodischer Posthumanismus auch ist, ohne selbtkritische Skepsis, methodisch-sprachkritische Anthropozentrik und das Abarbeiten an humanistischen Idealen beißt sich die Katze in den Schwanz. KI-Ethik kann jedoch an der Reflexion dieser kritischen Einwände wachsen. Darum sollte differenziert zwischen den methodischen Chancen solcher Kritik und den dogmatischen Irrwegen unterschieden werden. Es geht dabei um die Wahrnehmung ethischer Probleme. Offensichtlich ist das Sichern menschlicher, politischer Autonomie viel wichtiger in Zeiten bedrohter Privatheit – durch Google oder Facebook – (Pauen/Welzer 2015, S. 273-286) als der energiezehrende Streit für autonome und gleichberechtigte – übrigens von Menschen hergestellte – technische

Güter. Gleiches gilt für das **permanente Ringen um Menschenrechte und Menschenwürde**, in deren ethischem Zentrum nun einmal Menschen stehen (Pollmann 2022). Hinzu treten Probleme durch weltweite Kriege und Umweltzerstörung, Beeinflussung demokratischer Wahlen durch Deep Fakes oder unabsichtliche, aber reale Diskriminierung durch intransparente Modellbildung in machine learning-Systemen.

*Unterlassung ist auch eine Handlung: Es ist darum eigens zu rechtfertigen, warum diese realen Probleme zugunsten posthumanistischer Gleichmacherei von Menschen und Maschinen hintangestellt werden sollten.*

Als Radialkraft, die unsere ethischen Kufen in der Bahn hält, wirkt wohl eine Vermittlung beider Sichtweisen: Ethik ist immer anthropozentrisch und KI ist nach wie vor schlicht von Menschen hergestellte „technische Informationsverarbeitung“. Jedoch hat Technik – einschließlich KI – auch so etwas wie „Ethik“ oder „Politik“, jedoch im übertragenen Sinne: sie verkörpert Werte, Normen, Menschen- und Weltbilder derjenigen, die sie hergestellt haben (Winner 1980). Ethische Wissensdiversität entfaltet sich in diesen reflexiven Wechselbewegungen. Daraus nun zu folgern, dass eine KI selbst ethisch handelt, kann zwar die zwischenmenschliche Debatte anfeuern oder vergiften. Jedenfalls steigt damit das Risiko zwischenmenschlicher Diskriminierung und des Rückfalls hinter Errungenschaften der Menschenrechte und Menschenwürde. Eine Rechtfertigung, diesen Preis zu zahlen und in eine selbstverschuldete Unmündigkeit zu fallen, steht aus ...

## Zusammenfassung

Vorliegender Beitrag schließt unmittelbar an meine Analysen zur Wissensdiversität (vgl. Funk 2024) an. Dabei wurden vier Spannungsfelder einschließlich – bildlich gesagt – Radialkräften der Ethik exemplarisch unterschieden (Abb. 1). Hierzu zählen das Ringen gegen

- I. die Vereinnahmung von Ethik durch industrielle Interessen,
- II. die Beschränkung auf top-down vorgegebene „Checklistenethik“,
- III. den lebensfremden fachphilosophischen Elfenbeinturm sowie
- IV. posthumanistische Irrtümer.

Es lassen sich von der Orientierung individueller Lebensentwürfe bis hin zur globalen Geostrategie zahlreiche Spannungsfelder ergänzen. Anhand von zwei Beispielen wurden die vier ausgewählten vertieft. Zuerst ging es um den AI Act und Ethikricht-

linien. Top-down-Regulierung läuft Gefahr, Ethik auf Guidelines zu reduzieren und dabei sowohl moralische Alltagspraxis, also auch Bottom-up-Entwicklungspotenziale zu übersehen. Das kann umso größere Probleme nach sich ziehen, wenn beim weltweiten „Export“ europäischer Richtlinien auch noch koloniale Fehler wiederholt werden. Industrielle Interessen zielen auf eine möglichst weiche, rechtlich bindende Regulierung zugunsten ethischer Selbstverpflichtungen ab. Ethik wird also gegen geltendes Recht ausgespielt. Auf der anderen Seite fürchten Unternehmen in Europa um Wettbewerbsnachteile, insbesondere wenn der KI-Markt ohnehin von global agierenden Firmen vor allem der US-Westküste beherrscht wird. Im AI Act wird das komplexe Problemfeld zu sehr auf Risikomanagement reduziert. Ethik ist mehr, sie ist auch prüfendes Vorausdenken darüber, was wir eigentlich im positiven Sinne von KI erwarten – also Debatte über den wünschenswerten „public benefit“ (Prainsack/Forgó 2024). Dafür bedarf es im authentischen Sinne des Wortes **ethisch-demokratischer Lebenskunst statt Multiplex-Choice-Bürokratie**.

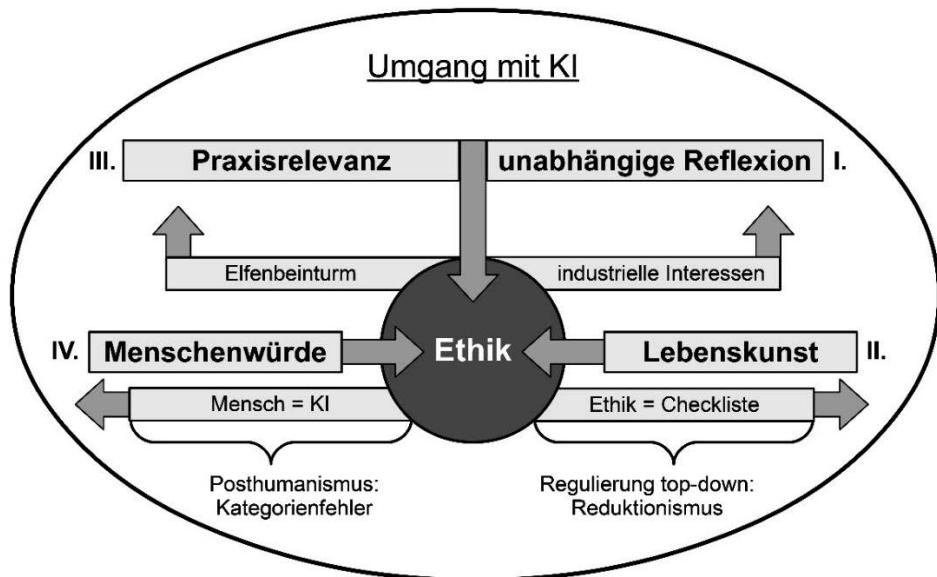


Abb. 1: Vier Radialkräfte der Ethik

Das zweite Beispiel rückte posthumanistische Kritik an menschenzentrierter, humanistischer Ethik in den Mittelpunkt. Dabei traten folgende Probleme auf:

1. Es gibt verschiedene Humanismen, Humanismus ist mehr als nur Anthroponozentrik, es gibt diverse Anthropozentriken (z.B. weltanschauliche oder methodische Anthropozentrik).
2. Physische Objekte einschließlich von Menschen hergestellter technischer Mittel werden nicht nur deskriptiv, sondern normativ als Alteritäten auf einer Stufe mit Menschen anerkannt.
  - ⇒ **Posthumanistischer Fehlschluss:** technische Informationsverarbeitung (KI) wird als moralischer Akteur auf eine Stufe mit Menschen gestellt; damit wird ein rationalistisch-reduktionistisches Weltbild aufgewertet (formale Sprache, propositionales Wissen, Zweck-Mittel-rational in von Menschen hergestellte technische Güter implementiert), das doch eigentlich als „anthropozentrischer Humanismus“ an den Pranger gestellt und überwunden werden soll – denn mit Humanismus ist auch die Epoche moderner Natur- und Technikwissenschaften angesprochen.
  - ⇒ Dieser Kategorienfehler wirkt toxisch auf die kritische Infrastruktur wirksamer Wissensdiversität I.
3. Der Blick wird von realen zwischenmenschlichen Ungerechtigkeiten abgelenkt, weiterhin von Problemen politischer Teilhabe sowie Herausforderungen interkultureller Wahrnehmungen und respektvollen Dialogs.
4. Auflösung rationaler Debatten in relativistische Beliebigkeiten, Stigmatisierung von Wahrheitssuche (statt Öffnung hin zur Diversität verschiedener Wissensformen).
  - ⇒ **Verstärkung der Gefahren** von Fakes, Propaganda, Gefährdung von Menschenrechten.

Die Ursache liegt im Defizit differenzierter Humanismus- und Anthroponozentrikforschung. Stattdessen wird diese im pauschalen Handstreich verdammt. „Narzisstische Bettgeschichten“ (Barad) können nicht nur von Menschen erzählt werden, die als „humanistisch“ oder „anthropozentrisch“ bezeichnet werden, sondern auch von Menschen, die sich selbst „posthumanistisch“ nennen. Relationalismus braucht ein methodisch anthropozentrisches Fundament (einschließlich Sprachkritik; Funk 2022b, S. 23-52), um nicht unabsichtlich politisch toxische Diskriminieren oder Silencing zu befördern. Auf der anderen Seite können gerade die vielfältigen Stimmen feministischer, postkolonialer oder posthumanistischer Kritik einen wichtigen Beitrag zur (menschlichen) Wahrnehmung gelebter Wissensdiversität leisten und schlussendlich die kritischen Infrastrukturen demokratischer Gesellschaften stärken. Es bleibt ein Balanceakt auf glattem Eis ...

## Literatur

- Barad, Karen (2020): *Agentieller Realismus*. Berlin: Suhrkamp.
- Birnbacher, Dieter (2022): *Klimaethik*. Stuttgart: Reclam.
- Braidotti, Rosi (2014): *Posthumanismus*. Frankfurt a.M./New York: Campus.
- Bridle, James (2023): *Die unfassbare Vielfalt des Seins*. München: C.H. Beck.
- Coeckelbergh, Mark (2022): *The Political Philosophy of AI*. Cambridge u.a.: Polity.
- Crawford, Kate (2019): *Atlas of AI*. New Haven/London: YUP.
- Dignum, Virginia (2019): *Responsible artificial intelligence*. Cham: Springer.
- Eke, Damian; Wakunuma, Kutamo & Akintoye, Simisola (Hrsg.) (2023): *Responsible AI in Africa*. Cham: Springer.
- Eubanks, Virginia (2018): *Automating Inequality*. New York: St. Martin's.
- European Parliament (2024): AI Act <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [Stand vom 28-07-2024]
- European Union (2019): Ethics guidelines for trustworthy AI, published 8 April 2019 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Stand vom 28-07-2024].
- Fischer, Johannes (2022): *Die Zukunft der Ethik*. Tübingen: Mohr Siebeck.
- Foot, Philippa (2023): *Die Natur des Guten*. Berlin: Suhrkamp.
- Funk, Michael (2022a): *Roboter- und KI-Ethik*. Wiesbaden: Springer.
- Funk, Michael (2022b): *Angewandte Ethik und Technikbewertung*. Wiesbaden: Springer.
- Funk, Michael (2023): *Ethik künstlicher Intelligenz*. Wiesbaden: Springer.
- Funk, Michael (2024): Nichtwissen ist Macht! Künstliche Intelligenz und ihre diversen Ethiken. In: Beinsteiner, Andreas; Dittrich, Ann-Kathrin & Hug, Theo (Hrsg.): *Medien – Wissen – Bildung: Wissensdiversität und formatierte Bildungsräume*. Innsbruck: iup, S. 41–58.
- Gardner, Howard (2002): *Intelligenzen. Die Vielfalt des menschlichen Geistes*. Stuttgart: Klett-Cotta.
- Gerdes, Anne (2022): The tech industry hijacking of the AI ethics research agenda and why we should reclaim it. *Discov Artif Intell* 2:25.
- Grunwald, Armin (Hrsg.) (1999): *Rationale Technikfolgenbeurteilung. Konzeption und methodische Grundlagen*. Heidelberg: Springer.
- Grunwald, Armin (Hrsg.) (2003): *Technikgestaltung zwischen Wunsch und Wirklichkeit*. Heidelberg: Springer.
- Grunwald, Armin (2017): *Technik und Politikberatung*. Frankfurt a.M.: Suhrkamp.
- Haraway, Donna (2018): *Unruhig bleiben*. Frankfurt a.M./New York: Campus.
- Hoerster, Norbert (2022): *Was ist Moral?* Stuttgart: Reclam.
- Höffe, Otfried (2014): *Die Macht der Moral im 21. Jahrhundert*. München: C.H. Beck.
- Hoppe, Katharina & Lemke, Thomas (2021): *Neue Materialismen*. Hamburg: Junius.

- Horster, Detlef (2009): *Ethik*. Stuttgart: Reclam.
- Hossenfelder, Malte (Hrsg.) (2013): *Antike Glückslehren*. Stuttgart: Kröner.
- Hubatschke, Christoph (2024): *Minoritäre Technologien*. Frankfurt a.M./New York: Campus.
- Ihde, Don (1993): *Postphenomenology. Essays in the Postmodern Context*. Evanston, Illinois: Northwestern University Press.
- Jonsen, Albert & Toulmin Stephen (1988): *The abuse of casuistry*. Berkeley: University of Chicago Press.
- Kallhoff, Angela (2022): *Der Mensch – das moralische Tier*. Berlin: Suhrkamp.
- Lenk, Hans (1982): *Zur Sozialphilosophie der Technik*. Frankfurt a.M.: Suhrkamp.
- Lenk, Hans (1994): *Macht und Machbarkeit der Technik*. Stuttgart: Reclam.
- Lenk, Hans (2007): *Global TechnoScience and Responsibility*. Berlin: LIT.
- Masek, Michaela (2023): *Antike Glücksethik*. Wien: facultas.
- Mbembe, Achille (2021): *Kritik der schwarzen Vernunft*. Berlin: Suhrkamp.
- Metzinger, Thomas (2019): Nehmt der Industrie die Ethik weg! Gastbeitrag im *Tagesspiegel* vom 08.04.2019 <https://www.tagesspiegel.de/politik/nehmt-der-industrie-die-ethik-weg-5548855.html> [Stand vom 28-07-2024].
- Murdoch, Iris (2023): *Die Souveränität des Guten*. Berlin: Suhrkamp.
- Nowotny, Helga (2023): *Die KI sei mit euch*. Berlin: Matthes & Seitz.
- Nussbaum, Marta (2020): *Gerechtigkeit oder Das gute Leben*. Frankfurt a.M.: Suhrkamp.
- Pauen, Michael & Welzer, Harald (2015): *Autonomie. Eine Verteidigung*. Frankfurt a.M.: Fischer.
- Pollmann, Arnd (2022): *Menschenrechte und Menschenwürde*. Berlin: Suhrkamp.
- Prainsack, Barbara & Forgó, Nikolaus (2024): New AI regulation in the EU seeks to reduce risk without assessing public benefit. *Nat Med*, Comment 18 March 2024. <https://doi.org/10.1038/s41591-024-02874-2> [Stand vom 28-07-2024].
- Rees, Martin (2017): Die organische Intelligenz hat keine langfristige Zukunft. In: Brockman, John (Hrsg.): *Was sollen wir von Künstlicher Intelligenz halten?* Frankfurt a.M.: Fischer, S. 36–39.
- Roser, Dominic & Seidel, Christian (2015): *Ethik des Klimawandels*. Darmstadt: WBG.
- Schmid, Thomas; Hildesheim, Wolfgang & Holoyad, Taras (2023): *Künstliche Intelligenz managen und verstehen*. Berlin u. a.: Beuth.
- Schwemmer, Oswald (1980): *Philosophie der Praxis*. Frankfurt a.M.: Suhrkamp.
- Sokolov, Daniel AJ (2024): UN-Generalversammlung: Mehr KI für Gutes, keine KI für Schlechtes, *heise* vom 22.03.2024 <https://heise.de/-9662662> [Stand vom 28-07-2024].
- Somek, Alexander (2021): *Moral als Bosheit*. Tübingen: Mohr Siebeck.
- Thurnherr, Urs (2010): *Angewandte Ethik*. Hamburg: Junius.
- Tsing, Anna (2022): *Der Pilz am Ende der Welt*. Berlin: Matthes & Seitz.
- Tugendhat, Ernst (1984): *Probleme der Ethik*. Stuttgart: Reclam.

- Unesco (2022): Recommendations on the Ethics of Artificial Intelligence. Adopted on 23 November 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi> [Stand vom 28-07-2024].
- United Nations (2024): General Assembly adopts landmark resolution on artificial intelligence, 21 March 2024 <https://news.un.org/en/story/2024/03/1147831> [Stand vom 28-07-2024].
- Williams, Bernard (1986): *Der Begriff der Moral*. Stuttgart: Reclam.
- Winner, Langdon (1980): Do Artifacts Have Politics? *Daedalus* 109 (1), S. 121–136.



# Notizen zum KI-Klima

**Oliver Leistert**

## *Zusammenfassung*

Mit KI ist eine Technologie zur Vermassung angetreten, die überwiegend zur Verschärfung sozialer Ungleichheiten und ökologischer Probleme beiträgt. Sie verstärkt den Klimawandel und macht uns blind gegenüber den Ursachen dieser Katastrophe. KI bewirkt eine Verlängerung bestehender Verhältnisse, die uns in diese tödliche Lage gebracht haben. Es muss deshalb darum gehen, gegen ihre Affirmation und Durchsetzung zu argumentieren.

## **Das Klima bestimmt unsere Lage**

Es lohnt sich, die Texte der KlimaforscherInnen zu lesen. Aber besser tagsüber und nicht kurz vor dem Schlaf. Aufgrund einer nächtlichen Lektüre des Textes „Klima und Wetter bei 3 Grad mehr“ von Stefan Rahmstorf war an diesen in jener Nacht nicht mehr zu denken. Dort las ich: „Die mittlere Erwärmung aller Landgebiete lag 2020 bei 2,0 Grad Celsius. Bei 3 Grad globaler Erwärmung sind bei uns also rund 6 Grad Erwärmung zu erwarten“ (Rahmstorf 2022, S. 14).

Während noch heute die Zahl von 1,5 Grad Celsius durch die Medien geistert, hatten wir es auf den Landflächen 2020 im Mittel bereits mit 2 Grad Celsius Erwärmung zu tun. Und die Chancen stehen nicht schlecht, dass 6 Grad im Mittel zu Lande bereits in wenigen Jahrzehnten auftreten werden. Im Mittel heißt aber auch, dass es immer mal deutlich mehr sein wird.

Wer behauptet, zu wissen, was das bedeutet, ist pathetisch. Ebenso fehlt es noch der komplexesten Computersimulation an epistemischen und ästhetischen Mitteln, diese

beinahe gesichert eintretende Katastrophe sinnlich begreifbar und somit vorstellbar zu machen. Sinnvoll ist deshalb das Eintreten auch für nicht-wissenschaftliche Zugänge zur sinnlichen Wahrnehmung des Klimawandels. Dies umfasst auch emotionale Zugänge (Schneider 2023; 2018).

Der Klimawandel ist nur eine Seite der Katastrophe. Der Verlust der Biodiversität ist die andere. Hauptgrund des Verlustes der Biodiversität ist die Zurichtung des Planeten durch extraktive und toxische Industrien. Die Erderwärmung beschleunigt das Artensterben, wenn sie z.B. durch Hitzeextreme Schwarze Flughunde in Australien zu Tausenden sterben lässt. Sie fallen einfach tot von den Bäumen (Kegel 2022, S. 46). Anders die Brüllaffen in Mexiko: sie fallen vor Schwäche von den Bäumen und sterben erst beim Aufprall (Boese & Demmer 2024).

Der menschengemachte Klimawandel und der menschengemachte Verlust an Biodiversität treten gemeinsam auf, haben beide ihren Grund in den fossilen Industrien, ihre Messbarkeit aber ist jeweils grundverschieden. Während der Klimawandel in nie bekannten Ausmaßen gemessen wird, vom All, unter dem Polar-Eis, mit tausenden Messstationen, passiert das Artensterben geradezu unbemerkt (Kolbert 2021). Viele Arten hatte der Mensch noch gar nicht registriert, da waren sie schon ausgestorben. Nur: mit der rapide schwindenden Biodiversität verschwindet auch die Basis des Lebens generell. Darin drückt sich vielleicht die Hybris der westlichen Gesellschaften am stärksten aus: zu glauben, „die Natur“ beherrschen zu können, wo es doch eine Dezimierung der Natur ist. Die Kehrseite hierbei ist im westlichen Jargon der Begriff der Ökologie, der nun, wo es für zu viele zu spät ist, in aller Munde ist: wir hängen alle zusammen drin. Allerdings aus sehr unterschiedlichen Ausgangspositionen und mit sehr unterschiedlichen Mitteln ausgestattet.

Ein viel zitierter Aufsatz der Klima- und Erdwissenschaften, der mehrfach aktualisiert und erweitert wurde, hat neun Prozesse definiert, die kritisch sind, um die Stabilität des „Systems Erde“ als Ganzes für den Menschen langfristig in ausreichendem Maße aufrechtzuerhalten (Richardson et al. 2023). Inzwischen wurden die planetarischen Grenzen von sechs dieser kritischen Prozesse überschritten, z.T. für sehr lange Zeiträume. Dies betrifft u.a. die Kohlendioxid-Konzentration in der Luft, die Integrität der Biosphäre und den Eintritt sog. „novel entities“ in die Umwelt. Dies sind von der chemischen Industrie hergestellte neuartige Stoffe, die z.T. für immer in der Umwelt bleiben werden (z.B. die meisten Plastiksorten). Recycling ist für die meisten dieser Stoffe mehr Propaganda als Realität. Unabhängige NGOs sagen, dass bisher nur 10% des bisher hergestellten Plastiks recycelt wurde. In vielen Ländern, z.B. den USA, fällt die Rate. Die Produktion von Plastik steigt global seit Jahren an (Swaminathan 2022). Plastik trägt

damit wesentlich zum Erhalt der Öl-Industrie bei und hängt direkt mit massiven CO2-Emissionen zusammen.<sup>1</sup>

Zurecht sind die Zweifel groß, ob es überhaupt einmal gelingen wird, die CO2-Emissionen global weit genug zu drücken, um an vielen Orten der Erde ein Überleben für eine ausreichend diverse Zahl an Lebewesen zu ermöglichen. Zu nachhaltig ist die Nicht-Nachhaltigkeit insbesondere westlicher Gesellschaften (Blühdorn et al. 2020). Auch ist die Reduktion von CO2 allein noch keine rettende Perspektive, vielmehr muss ein Ausstieg aus der modernen Lebensweise insgesamt vollzogen werden, oder wenigstens eine drastische Reform derselben, insbesondere um die Biodiversität einigermaßen zu erhalten (Beckert 2024). Der Ressourcenverbrauch muss runter. Das Artensterben muss verlangsamt werden. In der Realität nimmt der Ressourcenverbrauch jedoch immer weiter zu, ebenso wie das Artensterben.

Alle Vorschläge für einen wirkungsvollen strukturellen sozial-ökologischen Umbau, der auch die Produktion von Waren und unsere westliche Lebensweise einschließt, kann systemisch im Kapitalismus ohne sehr große Krisen nicht gelingen. Abkehr vom Wachstumsparadigma ist keine Option. Die strukturelle Blindheit des Kapitalismus gegenüber seinen eigenen sozialen und ökologischen Voraussetzungen wird die Lage weiterhin verschärfen. Ja, sogar „[d]ie Grenzen der liberalen Demokratie als eine dem Kapitalismus im globalen Norden lange Zeit adäquate politische Form zeichnen sich ab“ (Brand/Wissen 2024, S. 21).

Dass das hegemoniale Paradigma des Marktes zur Vermittlung gesellschaftlicher Prozesse, das uns in diese tödliche klimatische Lage insbesondere seit der Deregulierung der Märkte ab den 1970er gebracht hat, per Emissionshandel die Emissionen senken soll, ist Ausdruck einer vollkommen fehlgeleiteten instrumentellen Rationalität. Ebenso wie die Phantasie, das CO2 im Boden verpressen zu können. Auf die Erlösung durch Technik zu warten und zu argumentieren, in naher Zukunft werde es technische Lösungen für die Bekämpfung des Klimawandels geben, hat, wo es nicht offensichtlich der Ablenkung dient, eher religiöse Konnotationen als mit Einsicht zu tun. Insgesamt wird insofern zynischerweise folgerichtig vermehrt die Frage nach einer Anpassung an den Klimawandel gestellt, während Debatten und Maßnahmen zu dessen Reduktion erschreckenderweise rückgängig sind. Dabei lautet die eigentliche Frage, ob wir es selbst jemals schaffen werden, unsere imperiale Lebensweise (Brand/Wissen 2024, S. 31 ff.) abzulegen. Wer dies öffentlich und eindringlich unter dem Schlagwort „system change“

---

1 Jennifer Gabrys (2011, S. 82 ff.) beschreibt eindringlich, inwiefern Plastik notwendig für den Aufstieg von ICT war und ist.

fordert, erfährt jedoch zunehmend Repression und Einschüchterung (Schönherr 2024). Wie wir einen Weg finden, in den Ruinen des Kapitalismus zu überleben (Tsing 2017) und uns dabei mit dem Anfang einer neuen Welt auseinandersetzen (Schneider 2023), wird maßgeblich mitbestimmt davon, wie unsere sozio-technische Einbettung organisiert ist. Insofern sich weiterhin große Teile der Gesellschaft von den Anforderungen, die Klimawandel und Artensterben an uns stellen, angewidert abwenden, und es nur noch um die eigenen Privilegien und die Verteidigung von Partikularinteressen geht, steht eine noch kompliziertere Zersplitterung der Gesellschaft bevor, als sie bereits durch die Logiken von Social Media, smarten Geräten und generell ICT vorstrukturiert ist. Angefeuert durch irrwitzige und beängstigende Personen in der Politik, z.B. Donald Trump, werden populistisch Scheinprobleme vor die echten in den Diskurs geschoben und die politische Agenda in den westlichen repräsentativen Demokratien zunehmend von rechts oder rechts außen bestimmt. Anstatt für Klimagerechtigkeit einzutreten, indem z.B. wenigstens die massiven Emissionen der ganz Reichen besteuert werden, wird nur mehr Politik gegen die geringsten Emittenten von CO2 gemacht: vulnerable Menschen, die keine andere Möglichkeit sehen, als ihre Heimat zu verlassen, weil dort Kriege und gewaltsame Konflikte um Rohstoffe für den globalen Norden toben, weil bereits weit fortgeschrittene Effekte des vom Norden verursachten Klimawandels das Leben dort unmöglich machen, oder weil andere legitime Fluchtgründe vorliegen.

## KI intensiviert den Ressourcenverbrauch

Mitten in dieser für viele Lebewesen bereits tödlichen Superkatastrophe landet nun das prädiktive Paradigma der KI als Massenphänomen. Die Marktkapitalisierung des mit Abstand wichtigsten Herstellers passender Chips, Nividia, übersteigt erstmals die von Apple und Microsoft (Tagesschau 2024), eine über 42-fache Überbewertung der erwarteten Einnahmen der nächsten 12 Monate (Megaw/Acton 2024).<sup>2</sup> Traditionell ein Hersteller von Graphikkarten, deren Rechenstärke in einem ähnlichen Feld liegt wie dem, das heutige KI-Modelle bearbeiten, werden diese Spezialchips jetzt und wohl für die nächste Zeit Nvidia aus den Händen gerissen.

Analog zur Geschichte der CPUs liegt auch in der Versionsgeschichte der KI-Chips von Nvidia eine Steigerung der Leistung und Effizienz pro Iteration vor, entsprechend

---

2 Nur wenige Tage später fällt der Kurs innerhalb weniger Stunden. \$550 Milliarden Marktwert sind ausgeradiert (Bradshaw/Steer 2024). Nvidia rangiert nur mehr an dritter Stelle. Es stellt sich damit auch die Frage der Sinnhaftigkeit, solch dynamische und schnell getaktete Ereignisse in Texten wie diesem aufzugreifen. Sinnvoll ist es insofern, als das Ultravolatile dieser Märkte ein Ausdruck des KI-Klimas zu sein scheint.

wiederholt sich der Diskurs. „Mit dem aktuellen System ‚Grace Hopper‘ habe man zum Beispiel den Textrobother ChatGPT innerhalb von drei Monaten mit 8.000 Nvidia-Chips und einer Leistung von 15 Megawatt trainieren können, sagte Huang. ‚Blackwell‘ schaffe dies mit 2.000 Chips und vier Megawatt“, schreibt die Tagesschau unter Berufung auf den Nvidia-Chef Jensen Huang (Tagesschau 2024). Durch den Rebound-Effekt sind jedoch solch Verkündigungen als PR zu bewerten. Gerade im Feld der ICT hat sich gezeigt, dass Effizienzsteigerungen nicht zu geringeren Emissionen und Ressourcenverbrauch führen, sondern i.d.R. zur *Steigerung* von Komputationen (Santarius et al. 2016). Die hohe Nachfrage nach diesen Chips signalisiert, dass sich dieses Muster auch im Feld der KI einstellen wird.

Insgesamt wirkt der KI-Boom auch wie die Rettung einer Tech-Industrie, die zuletzt aufgrund global gestiegener Zinsen sowie bereits langanhaltender Einfallslosigkeit und kaum Erfolgen im Feld neuer Produkte eher am Kriseln als am Boomen war – Stichwort „Metaverse“. Die Zeiten, in denen Silicon Valley hoch angesehen ist, sind eher am Abklingen. Zu deutlich ist inzwischen geworden, wie schädlich deren Hegemonie für Gesellschaften und Umwelt ist. Diese Einsicht kommt allerdings zu spät. Längst haben sich die Tech-Konzerne strukturell und systemisch in alle Belange eingebettet.

Einmal mehr wird ein Technologie-Hype produziert, der, ähnlich wie wenige Jahre zuvor der Hype um Blockchains, verspricht, dass KI auch im Kampf gegen den Klimawandel hilfreich sei (Masterson 2024). Um dies zu bewirken, wächst nun zunächst einmal die Zahl der Datenzentren weltweit rasant an. Aktuellste Schätzungen gehen von derzeit rund 12.000 Datenzentren aus, die allein (ohne Infrastrukturen und Endgeräte) für rund 4% des globalen Energieverbrauchs verantwortlich sind. Die Nachfrage nach Rechenleistung ist größer, als es verfügbaren Strom gibt.<sup>3</sup> Das IT-industriefreundliche Irland, wohin die Schwergewichte aus dem Silicon Valley dank Niedrigsteuern ihren EU-Schwerpunkt gelegt haben, muss nun bereits 20% seines Stroms an Meta & Co abgeben (Campbell 2023). Durch KI steigen Strom- und Ressourcenverbrauch stärker an, und die damit verbundenen Naturzerstörungen sowie sozio-kulturellen Verwüstungen auch. Genaue Prognosen sind schwierig, aber lassen für die Zukunft keine Trendwende erkennen (De Vries 2023), denn bereits der globale Überwachungsapparat, der den KI-Boom mitvorantriebt, hat direkten Einfluss nicht nur auf die demokratischen, sondern

---

3 Dies sind ungefähre Angaben aus unterschiedlichen Quellen aus dem Internet, die hier nicht besonders ausgewiesen werden. Niemand scheint genau zu wissen, wie viele Datenzentren es gibt und wie viel Strom sie verbrauchen. Auch veralten die Zahlen sehr schnell, da das Feld sehr dynamisch ist. Es sind hunderte neue Rechenzentren weltweit geplant. Eine Übersicht über die Schätzungen für die Jahre 2007 bis 2021 liefern Mytton/Ashtine (2022). Auch ist zu beachten, dass Datenzentren immer größer werden. Die größten Komplexe können weit mehr als 1.000.000 m<sup>2</sup> umfassen.

auch auf die ökologischen Bedingungen vieler Menschen (Hogan 2015). Neben der weiterhin überwiegend fossilen und atomaren Erzeugung des Stroms zum Betrieb der geschätzt 12.000 Datenzentren weltweit werden riesige Mengen Wasser zur Kühlung dort benötigt, wo dies nicht mit Strom passiert. Wasser wird in weiten Teilen der Welt vermehrt zu einer knappen Ressource. Wieviel davon zur Kühlung der Datenzentren verwendet wird, ist unbekannt (Ristic et al. 2015) und wird von der Industrie strategisch verschwiegen.<sup>4</sup>

Angesichts dieser Entwicklungen gibt es Diskussionen, ob die Rechnerei selbst, aufgrund ihrer katastrophalen Ökobilanzen, juristisch als Ökozid aufgefasst werden sollte (Comber/Eriksson 2023). Die komplexen, teils neokolonialen und extrem extraktiven Lieferketten von (KI-)Chips, deren hochenergetische, multi-toxische Herstellung bei sehr kurzer Lebensdauer (meist unter drei Jahren) und ungelöstem Recycling (Cubitt 2017; Gabrys 2011), stehen an der Spitze eines Produktionsexzesses der Spätmoderne: Nur Hand in Hand mit z.B. der Weiterentwicklung der chemischen Industrien wurde es möglich, die immer leistungsfähigeren Chips aus dem Grundstoff Silizium zu ätzen und Reinheitsgrade zu erhöhen. 99% des Materials im Herstellungsprozess ist Abfall und überwiegend giftig (Gabrys 2011, S. 26).

Das Wissen, dass die Cloud gar keine ist, sondern auf Materie und Energieverbrauch beruht und dabei auch für gravierende Lärmverschmutzung sorgt, da sie sich auch gern in Wohngebieten niederlässt (Monserrate 2022), ist im täglich gelebten Exzess kaum reflektierbar, geschweige denn in einen reflektierten Gebrauch des Internet übersetzbare. Zu lange wurde der Unsinn verbreitet, dass es sich beim Internet um etwas Immaterielles handle.<sup>5</sup> So schön, richtig und intellektuell überzeugend Konzepte wie „digitale Suffizienz“ (Santarius et al. 2023) oder andere, gerechtere Arten der Ermittlung des ökologischen Fußabtrucks (Pasek et al. 2023) auch sind, bleiben sie doch konzeptionelle Übungen, solange das Kapital unbegrenzt den digitalen Betrieb bestimmt (Grünberger 2023). Es handelt sich letztlich um einen historisch gewachsenen und systemischen Zusammenhang von Kommunikation, Kohlenstoff und Kapitalismus (Brevini/Murdock 2017). Die Verschränkung der drei ist die Trajektorie, auf der die Abstraktionen moderner Vergesellschaftung mittels Extraktionen seit Eisenbahn und Telegrafen Raum und Zeit rekonfigurieren. Dies gilt uneingeschränkt auch für KI, die auf fortgesetztem Extraktivismus aufbaut: „AI is already a seamless part of a wider system where extraction

4 Eine recht gute Übersicht über die opake Lage gibt Gendries (2023). Das Problem soll hier nur anmoderiert werden.

5 In dieser Hinsicht ist die Ideologie des immateriellen Internet der Religion verwandt, die auf das Spirituelle setzt.

follows closely on the heels of abstraction, where everything of the world is seen as a utilitarian resource, not as a component of a fragile ecosystem“ (McQuillan 2022, S. 100).

Es ist unbestritten, dass KI diverse Prozesse positiv unterstützen kann, wie z.B. Wetterprognosen (Edwards 2013). Klassisch auch der Verweis auf die Medizin. Ähnlich werden auch kapitalintensive genverändernde Biotechnologien legitimiert. Leider wirkt KI jedoch in den allermeisten Anwendungsfällen, die auf soziale Fragen antworten sollen, als Verstärker bereits fataler oder ungerechter Dynamiken. *ChatGPT* als populistisches Interface zur KI ist dabei nur die offensichtlichste Seite des „power grab“ (Liu 2018). Verdeckter, aber umso effektiver ist dies insbesondere dort der Fall, wo durch lange Jahre neoliberaler Politiken Austerität und Prekarität zum Modell gesellschaftlicher Organisation geworden sind. Die Anwendung von KI in Institutionen und Unternehmen führt vermehrt zur Aushöhlung von Verantwortung, zu vermehrter Diskriminierung und insgesamt in der Tendenz zu autoritären Gefügen. Hier ließe sich einiges anführen: die als Objektivität verkaufte Reduktion, die KIs auszeichnet, ihre Gleichgültigkeit gegenüber Kausalitäten, ihre Imitation von Wissenschaft, und vollends ihre Automatisierung von Ungleichheit in sozialstaatlichen Systemen. Dies zu erkennen ist nicht leicht, da ihr gar nicht objektives Wirken schwer zu enttarnen ist: „AI acts as an epistemic power grab that conceals politics and ideology under its machinic opacity“ (McQuillan 2022, S. 51). Die epistemische Gewalt der KI schwächt andere, vielschichtigere und plurale Zugänge zur Welt. Auf genau solche, z.B. situiertes Wissen und lokale Erfahrungen, sind wir jedoch angesichts der Klimakatastrophe angewiesen.

## **Smartness und KI immunisieren Systeme gegen generische Krisen**

Die Einbindung von KI in den ICT-Komplex passiert keineswegs isoliert. Als eine Technologie der Prädiktion kann sie in vielen Bereichen eingesetzt werden, egal ob sie nun das nächste Wort errechnet und ausgibt oder die nächste Tat errechnet und exekutiert. Es geht daher um den Gesamtverbund „smarter“ Technologien und einer spezifischen, damit verbundenen Vorstellung vom Management der Gesellschaft.

Gesellschaft ist dabei begrifflich in den Kontext von „Smartness“ nicht übersetzbare. Anstelle eines Begriffes von Gesellschaft werden vielmehr Populationen verrechnet. Das „Smartness Mandate“ (Halpern et al. 2017) organisiert über geographisch abstrahierte Zonen das Medium Population im Sinne einer ewigen Iteration zwecks Optimierung. Mittels selbst-organisierender Systeme, die durch permanente algorithmische Traktierung massiver Datenströme aller Art ihre Subsysteme dezentral optimieren, sollen Resi-

lienzen gegen generische Krisen erwachsen. Die Blindheit solcher Systeme gegenüber den Ursachen und Spezifika einer Krise schließt jede soziale Frage von vorneherein aus. Es geht nicht um Ursachenbekämpfung oder überhaupt nur um ein Verständnis von Ursachen für bestehende Krisen. Das „Smartness Mandate“ kalibriert ohne externen Eingriff, d.h. immanent, in einer Abfolge unendlicher experimenteller Versionen unsere Existenzbedingungen, so die Vorstellung, die Orit Halpern et al. (2017) diskutieren.

Es gibt hierbei kein Kriterium für Erfolg oder Misserfolg, keine normative Basis, die zur Beurteilung der Lage herangezogen werden könnte. Ähnlich wie bereits von Gilles Deleuze einst für die Kontrollgesellschaften skizziert (Deleuze 1992), werden Elemente immer wieder neu moduliert. Andauernde Upgrades sorgen für stets neue Testphasen, die nie enden, da die Systeme sich selbst steuern und sich anhand neuer Konfigurationen durch die Zeit hängeln. Populationen bestehen dabei nicht aus Menschen, sondern aus deren und weiteren Daten, die nach biologischen Gesichtspunkten der Populationsforschung bearbeitet werden. Datenpopulationen können beliebig umgruppiert, unterschieden und je nach Anforderung optimiert werden. Dies beinhaltet einerseits biopolitische Aspekte, andererseits, da es sich um Datenpopulationen handelt, und da der Gesamtrahmen von Unternehmen und weniger von staatlicher Seite gestaltet wird, stehen Konsumverhalten und die dahinter stehenden Logiken der Logistik im Mittelpunkt des Interesses. Optimierung aller Prozesse ist das endlose Ziel, das den Rechnerverbund und dessen ständige Expansion rechtfertigt. Kein Bereich kommt mehr ohne algorithmische Optimierung aus: „Shopping, dating, exercising, the practice of science, the distribution of resources for public schools, the fight against terrorism, the calculation of carbon offsets and credits“ (Halpern et al. 2017, S. 119).

All dies legitimiert implizit, dass die Verdatung immer weiterer Bereiche immer granulärer voranschreitet. Das System läuft nur deshalb noch suboptimal, so die Auskunft, weil die Datenlage noch nicht hinreicht. Dies ist ein in jeglicher Hinsicht fatales, erneut exzessives Dispositiv, schon allein, da es sich zu den planetarischen Grenzen gegenläufig verhält. Das Ziel, das epistemisch in diesem Verbund ausschließlich möglich ist, kann nur die Verbesserung der Resilienz des Systems selbst sein, d.h. die Härtung des Systems gegenüber äußeren Krisen. Es passt sich dynamisch an eine sich plötzlich und kontingent ändernde Umwelt an. Diese ist keine biologische Umwelt, wie sie sich C.S. Holling mit dem Konzept der Resilienz einst vorstellte (Holling 1973), sondern, erneut, eine Abstraktion: „smartness abstracts the concept of resilience [...] into an all-purpose epistemology and value, positing resilience as a more general strategy for managing perpetual uncertainty and encouraging the premise that the world is indeed so complex that unexpected events are the norm.“ (Halpern et al. 2017, S. 122) Zum Repertoire gehört dann folgerichtig auch, dass suboptimale Populationsgruppen für den System-

erhalt aufgegeben werden können. KI fügt sich hier nahtlos ein: „Socially applied AI is ultimately a technology of unfreedom because it closes off possible futures other than those of its own determination“ (McQuillan 2022, S. 82).

In einer derart abstrahiert verwalteten Welt, die die Simulation biologisch-evolutionärer Prozesse als Paradigma des Sozio-Kulturellen behandelt (und somit auch kein historisches Motiv kennt), sind politischer Diskurs und Aushandlung einerseits zu Gunsten maschineller Modulationen entmachtet, andererseits kehren beide in veränderter, negativer Gestalt wieder zurück: da die Frage der Ursachen gesamtgesellschaftlicher Probleme vom System aussortiert wurde, mutieren solche Frage nun zu Figuren, bei denen Sprache und Aushandlung zunehmend durch Affekt, Rohheit und Gewalt sublimiert werden oder bereits ersetzt sind. Eine vermehrt algorithmisch abgetastete und medientechnologisch abstrahierte Gesellschaft verliert und verlernt, dies lässt sich leider an vielen konfliktuellen Topoi beobachten, basalste Techniken des Zusammenlebens, wie Empathie, Fürsorge und einen würdevollen Umgang miteinander. Die Kehrseite zunehmend algorithmisch vermittelter Erlebnisse und Alltagssituationen, in denen die Möglichkeit der Erfahrbarkeit unserer gegenseitigen Abhängigkeit durch opake, „smarte“ Systeme verstellt ist, ist der Rückzug des Sozialen (Andrejevic 2022), d.h. aber gleichzeitig auch der Niedergang von alternativen gesellschaftlichen Modellen.

## Gegen jede KI-Affirmation

Wollen wir die „carelessness towards perspective and standpoint“ (McQuillan 2022, S. 13) der KI überwinden, so hilft ein Hinweis von Isabelle Stengers: „It is not a matter of an observation of impotence, but rather of a point of departure“ (Stengers 2015, S. 30). Dieser könnte z.B. damit beginnen, konsequent die Aussage zurückzuweisen, dass es eine höhere, größere oder bessere Intelligenz als den Menschen durch KI gibt oder geben wird. Denn hier schließen nahtlos menschenverachtende Ideologien an, die zwischen Wertigkeiten von Leben unterscheiden. Auch sind wissenschaftliche Publikationen zurückzuweisen, die die Überlegenheit der KI gegenüber dem Menschen behaupten. Beispielhaft sei ein Text erwähnt, der behauptet, dass KI gegenüber dem Menschen im Feld der Kulturtechniken Schreiben und Illustration weniger Emissionen produziere (Tomlinson et al. 2024).

Solange Technologien die Leerstelle des westlichen Denkens markieren (Stiegler 2009), bzw. sogar der Begriff selbst ungeklärt ist (Schatzberg 2018), wird auch der Exzess der KI weitergehen können, diskursiv wie materiell. Obwohl der Planet und sein Klima von Technologien instrumenteller Rationalitäten bereits arg zugerichtet und irreversibel

kritisch verändert wurde, suchen externalisierende Gesellschaften weiterhin die Lösung ihrer Probleme in den Mitteln, die die Ursachen ihrer Probleme darstellen.

Wo ein struktureller Umbau aller Sphären mit dem Ziel einer wirkmächtigen sozial-ökologischen Wende nicht vorankommt, kann insbesondere an einer Stelle dennoch mit Erfolg gearbeitet werden: es geht darum, ein kritisches Nachdenken über Technologien in allen MINT-Studiengängen zu fördern und fest im Curriculum zu verankern. Wenn die Menschen, die diese tödlichen Technologien konzipieren oder weiterentwickeln, eine die Wirkungen ihres Schaffens reflektierende Ausbildung erhalten und bereits im Studium zwischen tödlichen und lebendigen Technologien zu unterscheiden lernen (Monnin 2023), wird ein Weg möglich gemacht, den Klimawandel langfristig aufzuhalten *und* Biodiversität zu erhalten. Durch extraktive, toxische, CO<sub>2</sub> emittierende und antisoziale Technologien, wie KI eine ist, wird das Leid hingegen nur verlängert werden.

Derweil steigen die Emissionen global weiter an (Millard/Murray 2024).

## Literatur

- Andrejevic, Mark (2022): Meta-Surveillance in the Digital Enclosure. *Surveillance & Society*, 20 (4), S. 390–396. <https://doi.org/10.24908/ss.v20i4.16008>,
- Beckert, Jens (2024): *Verkaufte Zukunft: Warum der Kampf gegen den Klimawandel zu scheitern droht*. Berlin: Suhrkamp.
- Blühdorn, Ingolfur; Butzlaff, Felix; Deflorian, Michael; Hausknost, Daniel & Mock, Mirjam (2020): *Nachhaltige Nicht-Nachhaltigkeit: Warum die ökologische Transformation der Gesellschaft nicht stattfindet*. Bielefeld: Transcript.
- Boese, Marie-Kristin & Demmer, Anne (2024): *Hitzewelle in Mexiko: Brüllaffen fallen tot von Bäumen*. tagesschau.de. 15. Juni 2024. Abgerufen unter: <https://www.tagesschau.de/wissen/klima/bruellaffen-hitze-artensterben-klimawandel-100.html> [Stand vom 19-06-2024].
- Bradshaw, Tim & Steer, George (2024): Nvidia share slide erases more than \$550bn in market value. *Financial Times*. 24. Juni 2024. Abgerufen unter: <https://www.ft.com/content/7332b1f8-cf7c-4bfa-82f4-88d0deb23f98> [Stand vom 14-06-2024].
- Brand, Ulrich & Wissen, Markus (2024): *Kapitalismus am Limit: öko-imperiale Spannungen, umkämpfte Krisenpolitik und solidarische Perspektiven*. München: oekom.
- Brevini, Benedetta & Murdock, Graham (Hrsg.) (2017): *Carbon Capitalism and Communication: Confronting Climate Crisis*. Cham: Springer International Publishing.

- Campbell, John (2023): Data centres use almost a fifth of electricity in Republic of Ireland in 2022. *BBC News*. 12. Juni 2023. Abgerufen unter: <https://www.bbc.com/news/articles/cpe9l5ke5jvo> [Stand vom 14\_06-2024].
- Comber, Rob & Eriksson, Elina (2023): Computing as Ecocide. *Ninth Computing within Limits 2023*. <https://doi.org/10.21428/bf6fb269.9fcdd0c0>.
- Cubitt, Sean (2017): *Finite media: environmental implications of digital technologies*. Durham: Duke University Press.
- De Vries, Alex (2023): The growing energy footprint of artificial intelligence. *Joule*, 7 (10), S. 2191–2194. <https://doi.org/10.1016/j.joule.2023.09.004>.
- Deleuze, Gilles (1992): Postscript on the Societies of Control. *October*, 59, S. 3–7.
- Edwards, Paul N. (2013): *A vast machine: computer models, climate data, and the politics of global warming*. Cambridge, Massachusetts: The MIT Press.
- Gabrys, Jennifer (2011): *Digital rubbish: a natural history of electronics*. Ann Arbor: University of Michigan Press.
- Gendries, Siegfried (2023): Rechenzentren und Wasser: Was Datennutzung mit Wasserkonflikten zu tun hat. 24. April 2023. *LebensraumWasser: Der Wasser-Blog*. Abgerufen unter: <https://www.lebensraumwasser.com/rechenzentren-und-wasser-was-datennutzung-mit-wasserkonflikten-zu-tun-hat/> [Stand vom 14\_06-2024].
- Grünberger, Nina (2023): Suffizienz, Digitalität und digitaler Kapitalismus. Herausforderungen für die Medienpädagogik. In: Beinsteiner, Andreas; Grünberger, Nina; Hug, Theo & Kapelari, Suzanne (Hrsg.): *Ökologische Krisen und Ökologien der Kritik*. Innsbruck: Innsbruck University Press, S. 109–124.
- Halpern, Orit; Mitchell, Robert & Geoghegan, Bernard Dionysius (2017): The Smartness Mandate: Notes toward a Critique. *Grey Room* 68, S. 106–129. [https://doi.org/10.1162/GREY\\_a\\_00221](https://doi.org/10.1162/GREY_a_00221).
- Hogan, Mél (2015): Data flows and water woes: The Utah Data Center. *Big Data & Society*, 2 (2). <https://doi.org/10.1177/2053951715592429>.
- Holling, Crawford Stanley (1973): Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, 4, S. 1-23. <https://doi.org/10.1146/annurev.es.04.110173.000245>.
- Kegel, Bernhard (2022): Biodiversität am Kipppunkt? In: Wiegandt, Klaus (Hrsg.): *3 Grad mehr: ein Blick in die drohende Heisszeit und wie uns die Natur helfen kann, sie zu verhindern*. München: Oekom Verlag, S. 31–54.
- Kolbert, Elizabeth (2021): *Das sechste Sterben: wie der Mensch Naturgeschichte schreibt*. Berlin: Suhrkamp,
- Liu, Hin-Yan (2018): The power structure of artificial intelligence. *Law, Innovation and Technology*, 10, S. 197–229. <https://doi.org/10.1080/17579961.2018.1527480>.

- Masterson, Victoria (2024): 9 ways AI is helping tackle climate change. World Economic Forum. Abgerufen unter: <https://www.weforum.org/agenda/2024/02/ai-combat-climate-change/> [Stand vom 14\_06-2024].
- McQuillan, Dan (2022): *Resisting AI: an anti-fascist approach to artificial intelligence*. Bristol: Bristol University Press.
- Megaw, Nicholas & Acton, Michael (2024): Nvidia overtakes Apple as its market capitalisation surpasses \$3tn. *Financial Times*. 5. Juni 2024. Abgerufen unter: <https://www.ft.com/content/1ec5523a-80ec-47ba-9f71-8765b4ae4577> [Stand vom 14-06-2024].
- Millard, Rachel & Murray, Clara (2024): Energy emissions hit record high on rising fossil fuel demand, says report. *Financial Times*. 20. Juni 2024 Abgerufen unter: <https://www.ft.com/content/f0e1f4fa-bc5a-45e9-9257-871dae461e5d> [Stand vom 20-06-2024].
- Monnin, Alexandre (2023): Zombie-Technologien und negative Commons. In: Beinsteiner, Andreas; Grünberger, Nina; Hug, Theo & Kapelari, Suzanne (Hrsg.): *Ökologische Krisen und Ökologien der Kritik*. Innsbruck: Innsbruck University Press, S. 35–54.
- Monserrate, Steven Gonzalez (2022): The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage. In: *MIT Case Studies in Social and Ethical Responsibilities of Computing*. <https://doi.org/10.21428/2c646de5.031d4553>.
- Mytton, David & Ashtine, Masaō (2022): Sources of data center energy estimates: A comprehensive review. *Joule*, 6 (9), S. 2032–2056. <https://doi.org/10.1016/j.joule.2022.07.011>.
- Pasek, Anne; Vaughan, Hunter & Starosielski, Nicole (2023): The world wide web of carbon: Toward a relational footprinting of information and communications technology's climate impacts. *Big Data & Society*, 10 (1). <https://doi.org/10.1177/20539517231158994>.
- Rahmstorf, Stefan (2022): Klima und Wetter bei 3 Grad mehr. In: Wiegandt, Klaus (Hrsg.): *3 Grad mehr: ein Blick in die drohende Heisszeit und wie uns die Natur helfen kann, sie zu verhindern*. München: Oekom Verlag, S. 13–30.
- Richardson, Katherine; Steffen, Will; Lucht, Wolfgang; Bendtsen, Jørgen; Cornell, Sarah E.; Donges, Jonathan F.; Drücke, Markus; Fetzer, Ingo; Bala, Govindasamy; von Bloh, Werner; Feulner, Georg; Fiedler, Stephanie; Gerten, Dieter; Gleeson, Tom; Hofmann, Matthias; Huiskamp, Willem; Kummu, Matti; Mohan, Chinchu; Nogués-Bravo, David; Petri, Stefan; Porkka, Miina; Rahmstorf, Stefan; Schaphoff, Sibyll; Thonicke, Kirsten; Tobian, Arne; Virkki, Vili; Wang-Erlandsson, Lan; Weber, Lisa & Rockström, Johan (2023): Earth beyond six of nine planetary boundaries. *Science Advances*, 9 (37). <https://doi.org/10.1126/sciadv.adh2458>.
- Ristic, Bora; Madani, Kaveh; Makuch, Zen (2015): The Water Footprint of Data Centers. *Sustainability*, 7 (8), S. 11260–11284. <https://doi.org/10.3390/su70811260>.
- Santarius, Tilma; Bieser, Jan C. T.; Frick, Vivian; Höjer, Mattias; Gossen, Maike; Hilty, Lorenz M.; Kern, Eva; Pohl, Johanna; Rohde, Friederike & Lange, Steffen (2023): Digital sufficiency: conceptual considerations for ICTs on a finite planet. *Annals of Telecommunications*, 78 (5), S. 277–295. <https://doi.org/10.1007/s12243-022-00914-x>.

- Santarius, Tilman; Walnum, Hans Jakob & Aall, Carlo (Hrsg.) (2016): *Rethinking Climate and Energy Policies*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-38807-6>.
- Schatzberg, Eric (2018): *Technology: Critical History of a Concept*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226584027.001.0001>.
- Schneider, Birgit (2023): *Der Anfang einer neuen Welt: wie wir uns den Klimawandel erzählen, ohne zu verstummen*. Berlin: Matthes & Seitz Berlin.
- Schneider, Birgit (2018): *Klimabilder: eine Genealogie globaler Bildpolitiken von Klima und Klimawandel*. Berlin: Matthes & Seitz Berlin.
- Schönherr, Harff-Peter (2022): Druck auf Klimaschützer: Staatsschutz geht gegen Fridays vor. *Die Tageszeitung: taz*. 10. Januar 2022. Abgerufen unter: <https://taz.de/!5824934> [Stand vom 19-06-2024].
- Stengers, Isabelle (2015): *In Catastrophic Times – Resisting the Coming Barbarism*. London: Open Humanities Press.
- Stiegler, Bernard (2009): *Technik und Zeit: Der Fehler des Epimetheus*. Zürich, Berlin: Diaphanes.
- Swaminathan, Arjun (2022): *Circular Claims Fall Flat Again*. Greenpeace US. Abgerufen unter: [https://www.greenpeace.org/usa/wp-content/uploads/2022/10/GPUS\\_FinalReport\\_2022.pdf](https://www.greenpeace.org/usa/wp-content/uploads/2022/10/GPUS_FinalReport_2022.pdf) [Stand vom 14.06.2024].
- Tagesschau (2024): Nvidia überholt Microsoft und Apple als wertvollstes Unternehmen. *tagesschau.de*. 19. Juni 2024. Abgerufen unter: <https://www.tagesschau.de/wirtschaft/digitales/nvidia-wertvollstes-unternehmen-boerse-chiphersteller-100.html> [Stand vom 19-06-2024].
- Tomlinson, Bill; Black, Rebecca W.; Patterson, Donald J. & Torrance, Andrew W. (2024): The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports*, 14 (1), 3732. <https://doi.org/10.1038/s41598-024-54271-x>.
- Tsing, Anna Lowenhaupt (2017): *Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton: Princeton University Press.



# Desiderata der AI-Literacy-Diskurse

Theo Hug

## *Zusammenfassung*

Der Beitrag befasst sich mit Artificial Intelligence (AI) Literacy und kontemporären Entwicklungen, die auf die Förderung von menschlichen Kompetenzen im Umgang mit Künstlichen Intelligenzen (KI) in Bildungsinstitutionen abheben. Die rasante Verbreitung von KI-Anwendungen in vielen gesellschaftlichen Teilbereichen stellt auch die Bildungssysteme vor große Herausforderungen. In Ergänzung zu aktuellen pädagogischen Relevanzformeln, die mit Schlagworten wie ‚digitale Bildung‘ und ‚digitale Kompetenz‘ sowie ‚Digital Literacy‘ und ‚Futures Literacy‘ verknüpft sind, werden neuerdings Konzepte und Praktiken der AI Literacy international diskutiert. Ausgehend von konzeptionellen Vorüberlegungen zur Vielfalt von Literalitäten werden im Beitrag einige Desiderata aus der Sicht der Medienbildungsforschung zur Diskussion gestellt.

## Auftakt

In neuerer Zeit ist parallel zu den multiplen Krisendiagnosen<sup>1</sup> ein verstärktes Interesse an der Zukunft des Lernens und an Bildungsprozessen in Kulturen der Digitalität zu beobachten. Dies zeigt sich einerseits an den intensiv geführten Diskursen über Zukunftskompetenzen, Futures Literacy und Future Skills (vgl. exemplarisch Miller 2018; Ehlers 2020). Andererseits sind die vielgestaltigen Redeweisen von ‚digitaler Bildung‘, ‚digitalen Kompetenzen‘ oder ‚digitalem Unterricht‘ nicht nur in bildungspolitischen, sondern auch in wissenschaftlichen und pädagogischen Diskurszusammenhängen international bedeutsam geworden. Die begrifflichen Unschärfen, die mit der Pluralisierung

---

1 Vgl. häufig verwendete Stichworte wie Klimakrise, Demokratiekrisen, Sozialkrisen, Finanzkrisen, Wirtschaftskrisen, Ungleichheitskrisen, Energiekrisen, Migrationskrisen, Ernährungskrisen oder Biodiversitätskrisen.

von Literalitätskonzepten und deren Anwendung sowie mit der Zuschreibung des Adjektivs ‚digital‘ zu nicht-technischen Phänomenbereichen verknüpft sind, spielen dabei eine höchst ambivalente Rolle. Sie suggerieren auf den ersten Blick Möglichkeiten der bereichsübergreifenden Verständigung und der zukunftsorientierten Bildungsentwicklung an den Nahtstellen von pädagogisch-praktischen, organisationalen, politischen, wirtschaftlichen und wissenschaftlichen Tätigkeiten. Diese erweisen sich aber bei näherer Betrachtung als problematisch.

Vor allem dort, wo Aufgaben der Politikberatung, die Akquise von Fördermitteln oder akademische Selbstvermarktung im Vordergrund stehen, sind fragwürdige Redeweisen von Literacies, Skills und Kompetenzrahmen vielerorts zur Normalität geworden. Die Verständigung weist dabei angesichts der begrifflichen Unschärfen und der metaphorischen Ausdrucksformen häufig einen scheinbaren Charakter auf. Dies tut jedoch der Beförderung gesellschaftlicher Normalisierungsprozesse im Umgang mit Fragen der Digitalität und der Digitalisierung keinen Abbruch. Ansprüche der begrifflich-theoretischen Differenzierung, der Explikation stillschweigender Annahmen oder der Aufklärung über Machtinteressen, historische Medienkulturentwicklungen und alternative Innovationspfade spielen in den entsprechenden Aufgabenkontexten bestenfalls eine untergeordnete Rolle. Ungeachtet dessen können kritische Reflexionen auch in diesem Zusammenhang erheblich zur Erweiterung von Denk- und Handlungsperspektiven und zur Wahrnehmung von Entwicklungspotenzialen beitragen. Dies gilt nicht zuletzt für die Propagierung von AI Literacy, die neuerdings im Ensemble zukunftsrelevanter Pädagogisierungsformeln *sensu* Veith (2003) neben Digital Literacy, Data Literacy und Futures Literacy sowie neben digitalen Kompetenzen und digitaler Bildung international diskutiert wird (vgl. exemplarisch Burgsteiner/Kandlhofer/ Steinbauer 2016; Long/ Magerko 2020; Ng et al. 2021; Casal-Otero et al. 2023; Shah 2023, S. 39-43; Kreinsen/ Schulz 2023; Ng et al. 2023; Rütti-Joy/Winder/Biedermann 2023; Stolpe/Hallström 2024). Ausgehend von konzeptionellen Vorüberlegungen zur Vielfalt von Literalitäten werden im Beitrag einige Desiderata aus der Sicht der Medienbildungsforschung zur Diskussion gestellt.

## **Vorüberlegungen zum Literacy-Begriff und zur Vielfalt von Literalitäten**

Ähnlich wie Bildung ist auch Literacy ein ‚Drehtürbegriff‘, der in unterschiedlichsten wissenschaftlichen, politischen, technologischen und alltagsweltlichen Zusammenhängen figuriert. Während der Ausdruck ‚Literalität‘ im Deutschen vergleichsweise weniger oft verwendet wird, müsste ein *A to Z* der Literacies als gedrucktes Kompendium in sehr kurzen Abständen aktualisiert werden. Von AI Literacy bis Zoological Literacy werden

zahlreiche Komposita in verschiedenen Diskurszusammenhängen mehr oder weniger differenziert diskutiert und es kommen ständig neue hinzu.<sup>2</sup> Nicht selten haben dabei begriffspolitische oder anwendungsorientierte Motivierungen die Oberhand, während begriffsanalytische Motivierungen, die sowohl die Modalitäten der begrifflichen Verknüpfungen als auch kontrastierende Grundbedeutungen von Literacy (vgl. Street/Lefstein 2007; Olson/Torrance 2009) explizit in den Blick nehmen, eher selten vorkommen.

Was die Grundbedeutungen von Literacy betrifft, so kamen zu den älteren Auffassungen, die sich auf Konzeptionen und soziale Praktiken des Lesens und Schreibens in kulturellen Kontexten beziehen (vgl. Street 1984, S. 1), etliche begriffliche Erweiterungen und Kontrastierungen hinzu. Lankshear und Knobel (2007) haben eine Gruppierung der neueren Entwicklungen vorgeschlagen, bei der zwei zentrale Kriterien angewendet werden: Jene neuen Literalitäten, bei denen über konventionelle Begriffsauffassungen („conventional literacies“) hinaus neue technologische Entwicklungen („technical stuff“) eine Rolle spielen, fallen unter die „peripheren Fälle“ von Literacies. Wenn zusätzlich neue Aspekte der Haltung, Gesinnung oder Gewohnheit ins Spiel kommen („ethos stuff“), dann ist von „paradigmatischen Fällen“ von New Literacies in einem weiteren Sinne die Rede (vgl. Lankshear/Knobel 2007, S. 7).

Stordy (2015) hat auf der Basis von Streets (1984) Unterscheidung von „autonomous literacies“ als kognitive Fähigkeiten und „ideological literacies“ als soziale Praktiken sowie der Dreiteilung von Lankshear und Knobel (2007) in „conventional literacies“, „peripheral cases“ und „paradigm cases“ eine Rahmenstruktur zur Verortung und Analyse der Vielfalt neuer Literalitäten entwickelt. Dabei legt er ein besonderes Augenmerk auf digitale Technologien. Einerseits lassen sich damit viele der neueren Bindestrich-

---

2 Hier eine Beispielsammlung in alphabetischer Reihung: *academic literacy, adult literacy, advertising literacy, AI literacy, algorithmic literacy, art literacy, assessment literacy, attentional literacy, biological literacy, biomedia literacy, coding literacy, computer literacy, consumer literacy, copyright literacy, critical literacy, critical science literacy, cultural literacy, data literacy, data visualization literacy, depression literacy, digital literacy, diversity literacy, ecological literacy, emotional literacy, energy literacy, environmental literacy, family literacy, feminist literacy, film literacy, financial literacy, food literacy, futures literacy, game literacy, gender literacy, geographical literacy, global literacy, hacker literacy, health literacy, ICT literacy, information literacy, Internet literacy, legal literacy, liberation literacy, library literacy, ludoliteracy, mathematical literacy, media literacy, meta-literacy, multicultural literacy, multiliteracies, network literacy, new media literacy, news literacy, numerical literacy, ocean literacy, online literacy, oral literacy, pattern literacy, peace literacy, physical literacy, physics literacy, political literacy, post-literacy, posthuman literacy, psychic literacy, psychological literacy, radio literacy, scientific literacy, sexual literacy, situated literacies, social presence literacy, sport literacy, STEM literacy, stone age literacy, sustainability literacy, technological literacy, television literacy, transformative literacy, transliteracy, urban literacy, vertical literacy, visual literacy, Web literacy, weblog literacy, zoological literacy, etc.* Einige konzeptionelle Überlegungen zur Ökologie von Literalitäten und zu den Dynamiken der „literacification of everything“ wurden an anderer Stelle veröffentlicht (s. Hug 2019).

Literalitäten trefflich diskutieren, andererseits weist die Rahmenstruktur auch Limitationen auf, die im Text von Stordy (2015) unberücksichtigt bleiben. Er betont zwar, dass er das „literacy framework“ nur auf jene Typen von Literacies anwendet, die direkt oder indirekt mit digitalen Technologien in Verbindung stehen (ebd., S. 456). Die Verknüpfung von analogen und digitalen Dimensionen, Überlegungen zu postdigitalen Entwicklungen oder Erwägungen im Zusammenhang transversal vernetzter Mediensysteme kommen dabei aber nicht in den Blick. Was Aspekte der Zeitlichkeit betrifft, so hebt er hervor, dass die „peripheral literacies“ letztendlich „traditional literacies“ werden und dass die Rahmenstruktur im historischen Zusammenhang zu sehen ist: „The framework is best perceived as time dependent“ (ebd., S. 471). Dass vor allem die „peripheral literacies“ bisweilen auch den Charakter von Modeerscheinungen haben, wird nicht erwähnt. Auch andere Aspekte von Zeitlichkeit werden nicht reflektiert. Das betrifft beispielsweise Fragen der Eigenzeit von Lern- und Bildungsprozessen, der technologisch imprägnierten Beschleunigung von gesellschaftlichen Entwicklungen oder des Verhältnisses chronobiologischer, physikalischer, subjektiv erfahrener und historisch kultureller Zeitauffassungen.

Unabhängig davon, ob nun von Literacies und Skills oder von Lernen und Bildung die Rede ist, die Bedeutung angemessener Zeiten, Tempi, Pausen, Taktungen und Rhythmen ist in allen diesen Phänomenbereichen reflexionswürdig. Analoges gilt für Relationen des Hier und jetzt im Kontrast zum Dort und Dann. Hinzu kommt die Gleichzeitigkeit ungleichzeitiger Literalitätsbegriffe, die mit den kompositionsfreudigen Entwicklungen einhergehen, die sich durch vielgestaltige Analogien, figurative Übertragungen und begriffliche Weiterungen auszeichnen. In den Literalitätsdebatten sind nicht nur Grundbegriffe in Verwendung, die sich auf Konzeptionen und soziale Praktiken des Lesens und Schreibens in kulturellen Kontexten oder damit verknüpfte Abstrakta wie Alphabetisierung, Schriftkundigkeit, Schriftlichkeit und Belesenheit beziehen. Die Verwendungsweisen beziehen sich mitunter auch ganz allgemein auf Bildung, Gebildetheit, kommunikative Kompetenzen, semantisches Verstehen, Interpretationsfähigkeit, Wissensorganisation, Sinnproduktion, Zukunftsfähigkeit sowie Bedeutungszuschreibungen und Orientierungsleistungen aller Art. Semantische Überschneidungen und Unschärfen betreffen damit nicht nur immanente Aspekte der Kategorisierung in der Rahmenstruktur von Stordy (2015, S. 471), sondern auch Literalitätsbegriffe, die eine Erweiterung seiner Taxonomie nahelegen. Aus systematischer Sicht kommen weiters subsumptionslogische Annahmen hinzu, die sich auf die Verortung von Komposita wie Mathematical Literacy, Oral Literacy oder Visual Literacy in den Literacy-Diskursen beziehen.

Diese Annahmen erweisen sich bei näherer Betrachtung in zweifacher Hinsicht als problematisch. Einerseits bestehen bei der Interpretation von Komposita und ihren Bestand-

teilen generell große Spielräume und Mehrdeutigkeiten, die je nach Verwendungszusammenhang förderlich oder hinderlich sein können. Auch dann, wenn im Einzelfall speziellere oder vergleichsweise sehr eng gefasste Bedeutungen zur Routine werden, bleibt ein Reflexionsbedarf hinsichtlich der impliziten oder expliziten Bedeutungsverschiebungen, der Relevanz begriffspolitischer Aspekte sowie der erhellenden und verschleiernden Aspekte der metaphorischen Anwendungen. Andererseits stellt sich die Frage, inwieweit die vielgestaltigen Analogien und figurativen Übertragungen in den Literacy-Diskursen der Diversität von Wissensformen und -praktiken gerecht werden können. Wenn wir Phänomenbereichen wie Numeracy, Orality, Visuacy und Audability<sup>3</sup> in epistemologischen, kognitiven und sozio-kulturellen Hinsichten eine relative Eigenständigkeit zugestehen, dann erweisen sich nicht nur Taxonomien wie die von Stordy (2015) als problematisch, sondern auch andere Kategorisierungsversuche von Literacies und viele wie auch immer motivierte literalitätsbezogene Wortschatzerweiterungen generell (vgl. Hug 2023).

### **Kritische Überlegungen zu einigen Desiderata der AI Literacy Diskurse aus der Perspektive der Medienbildungsforschung**

In den öffentlichen Diskursen über Digitalisierung (in) der Bildung nehmen KI-Entwicklungen neuerdings eine herausragende Stellung ein. Die Relevanzbekundungen drücken mitunter eine Aufbruchsstimmung aus, die zur umfassenden Nutzung von KI-Potenzialen im Bildungswesen ermuntert:

„While we should be excited to embrace the full potential of AI, we need to be thoughtful and deliberate in our approach to ensure it aligns with our ethical principles.“ (Shah 2023, S. 208)

Für die Beachtung ethischer Dimensionen sind dabei die Lehrkräfte zuständig:

„Educators will need to build robust policies at the school and classroom levels to ensure that responsible and ethical use of AI is encouraged and required.“ (Shah 2023, S. 183)

In der Entfaltung der Umarmungs metaphorik wird von Shah (2023) allerdings nicht spezifiziert, wie angesichts der datenextraktivistischen Praktiken der globalen IT-Unternehmen und den vergleichsweise schwachen KI-Regulierungen in den unter-

---

3 Damit sind hier „acoustic abilities“ im Sinne von klangbezogenen Fähigkeiten und Kompetenzen gemeint.

schiedlichen Weltregionen die Lehrkräfte dafür Sorge tragen können, dass Anforderungen der Transparenz, Fairness, Privatsphäre, Humanität, Inklusion und des Datenschutzes angemessen Beachtung finden können.

Der Stellenwert von KI-Anwendungen in der Bildung wird gelegentlich auch im Rückgriff auf die Märchenwelt gerahmt. So beziehen sich beispielweise Seldon, Metcalf und Abidoye (2020) in ihrem Buch über eine vierte Bildungsrevolution, die in Analogie zu Industrie 4.0 mit dem Label „Education 4.0“ charakterisiert wird, auf das „Aschenputtel-Motiv“. Im Zusammenhang ihrer vorsichtig optimistischen Einschätzung zum Einsatz von KI und Robotern im Bildungswesen schreiben sie: „As with AI in general, education is the Cinderella subject of the vast robot literature“ (ebd., S. 152). Insgesamt geht es ihnen um eine effektivere Gestaltung von Bildungsprozessen und um eine rasche Wendung der Aschenputtel-Position von KI im Bildungswesen: „*AI should be the fairy-tale princess or the Prince Charming in education.*“ (ebd., S. 0 [sic!], kursiv i. Org.)

Solche Auf- und Umbruchstimmungen im Zusammenhang digitaler Transformationsprozesse drücken keineswegs einen Konsens aller gesellschaftlichen Gruppen, Milieus und Schichten aus. Im Gegenteil ist in den Bildungskulturen der Digitalität vielfach ein Unbehagen zu verzeichnen (vgl. Hug 2021). Die bildungspolitischen Forderungen nach der Förderung von AI Literacy auf globaler, EU-weiter oder nationaler Ebene (vgl. exemplarisch UNESCO 2022; EU 2024)<sup>4</sup> korrespondiert entsprechend nicht nur mit Ansprüchen aus diversen industriellen und wirtschaftlichen Sektoren, sondern auch mit der ambivalenten Stimmungslage und partiell auch mit aufklärerischen Motiven.

Die Frage, was unter AI Literacy konkret verstanden werden kann, wird dabei nicht einheitlich beantwortet. In den verfügbaren Literatur-Reviews zeichnet sich ein komplexes Bild unterschiedlicher Akzentuierungen ab. Long und Magerko (2020) definieren AI Literacy als ein „*set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace*“ (ebd., S. 2, kursiv i. Org.). Sie verorten AI Literacy im Umfeld von anderen Literacies<sup>5</sup> und sortieren 150 thematisch relevante Dokumente im Sinne einer „Scoping Study“ (ebd.) entlang von 17 Kompetenzen und 15 Design-Gesichtspunkten (s. Abb. 1).

4 Was die Situation Österreich betrifft, so fand am 3. Juli 2024 im Rahmen der Digitalen Kompetenzoffensive eine Kick-off Veranstaltung „KI-Literacy in Österreich: Standortbestimmung und Ausrichtung für eine digitale Zukunft“ statt (s. <https://www.digitalaustria.gv.at/WissensWert/Events-Digital-Austria/kick-off-ki-literacy.html>).

5 Das betrifft insbesondere Digital Literacy, Data Literacy, Computational Literacy und Scientific Literacy.



Abbildung 1: AI Literacy – Kompetenzen und Design-Gesichtspunkte nach Long und Magerko (2020)  
(Quelle: <https://aiunplugged.lmc.gatech.edu/ai-literacy/>)

Long und Magerko (2020) nehmen in ihrer Studie Fragen im Zusammenhang von „AI education for non-technical learners“ (ebd., S. 3) in den Fokus. Ungeachtet der knappen Erläuterungen wird dabei ein breites Spektrum von Kompetenz-Dimensionen und Design-Gesichtspunkten deutlich, in dem allerdings auch Desiderata auszumachen sind. Ähnlich verhält es sich mit der explorativen Studie von Ng et al. (2021), in der 30 beurachtete Beiträge im Hinblick auf definitorische, didaktische, evaluative und ethische Dimensionen von AI Literacy untersucht werden. Die Dokumente werden in dieser Studie auf dem Hintergrund eines Coding Frameworks, der Bloomschen Taxonomie sowie des *Technological, Pedagogical and Content Knowledge* (TPACK) Modells (Mishra/Koehler 2006) analysiert. Das TPACK-Modell kommt dabei in einer fachdidaktischen Interpretation mit Fokus auf Unterrichtsstrategien, Lehr-/Lernkonzepte, Classroom Management sowie adaptive und personalisierte Lernformen in einigen Bereichen der informatischen Bildung zur Anwendung. Alternative Interpretationen dieses Modells im Sinne eines Kompetenzmodells für eine fächerübergreifende, lernkulturorientierte oder etwa lebensweltorientierte Medienbildung (vgl. Schmid/Petko 2020) werden nicht berücksichtigt.

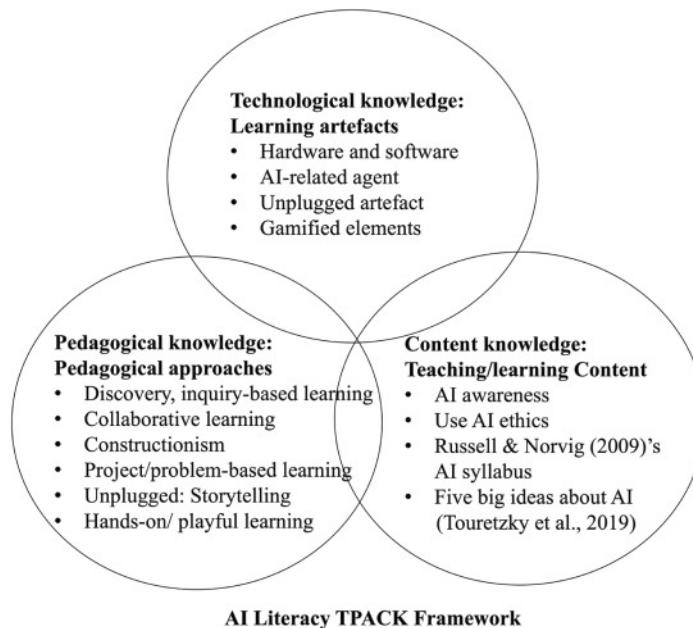


Abbildung 2: AI Literacy im TPACK-Modell (Quelle: Ng et al. 2021, S. 5)

In ihren abschließenden Empfehlungen für zukünftige „AI literacy education“ betonen Ng et al. (2021, S. 9-10), dass KI-Kompetenz im 21. Jahrhundert Teil der Grundbildung für alle sein und umfassend sowie altersgerecht vermittelt werden sollte. Studierende sollen aktiv Problemlösungen mit KI entwickeln und nicht nur als NutzerInnen agieren. Lehrkräfte müssen ihr KI-Wissen aktualisieren und passende pädagogische Rahmenwerke nutzen, um den Unterricht zu verbessern. Künftige Forschung sollte Lehrstrategien und Bewertungsmethoden entwickeln, die sowohl die technischen Fähigkeiten als auch das soziale Verantwortungsbewusstsein der Lernenden fördern. Was pädagogische und didaktische Perspektiven betrifft werden projektbasierte Lernformen, Strategien der Gamification und konstruktionistische Modelle hervorgehoben.

„Future researcher and educators will develop pedagogical strategies (e.g., collaborative project-based learning, gamification) and theoretical models (e.g., self-determination theory, constructionism) to increase students' motivation and engagement, promote interaction and collaboration, enhance motivation and attitudes, and develop numerous learning skills in the context of AI literacy.“ (Ng et al. 2021, S. 10)

Zu ähnlichen Schlussfolgerungen kommen auch Casal-Otero et al. (2023) in ihrem systematischen Literatur-Review von Beiträgen in der Datenbank Scopus zum Thema „AI literacy in K-12“. In ihrer Analyse von 179 ausgewählten Beiträgen aus den Jahren 2004 – 2023 identifizieren sie zwei große Gruppen von AI Literacy, von denen sich die erste auf Lernerfahrungen und die zweite auf Implementierungsvorschläge<sup>6</sup> bezieht. Sie betonen das Erfordernis eines „competency frameworks“ und einer klaren Definition von AI Literacy, ohne selbst im Anschluss an die analysierten Beiträge eine solche zu explizieren (vgl. ebd., S. 11). Was die Integration von AI Literacy in Curricula und Lehrpläne betrifft, so plädieren sie für einen interdisziplinären und kompetenzbasierten Ansatz und die Integration von KI-Kompetenzen in die diversen Fächer sowie für disziplinübergreifende praxisnahe Ansätze und projektbasierte Arbeitsformen. Hand in Hand damit soll Casal-Otero et al. (2023) zufolge AI Literacy zur Erweiterung des Lernens in den Fächern eingesetzt werden, wobei die aktive Beteiligung der Lehrkräfte in Kooperation mit PädagogInnen und AI ExpertInnen hohe Priorität hat (vgl. ebd., S. 12-13).

Die Vielzahl der Ansätze und Konzeptualisierungen korrespondiert mit einer Pluralität von Auffassungen hinsichtlich der Gewichtung relevanter Merkmale von AI Literacy. Eine gemeinhin akzeptierte Definition ist nicht in Sicht, wobei in den diversen Versuchen des Assessment und der Evaluierung von AI Literacy die oben erwähnte Definition von Long und Magerko (2020, S. 2) in der empirischen Forschung zumindest mitberücksichtigt wird (vgl. exemplarisch Carolus et al. 2023; Laupichler/Aster/Raupach 2023; Ehlers et al. 2024).

Als Zwischenfazit lässt sich festhalten, dass sich die skizzierten AI-Literacy-Diskurse ungeachtet der unterschiedlichen Ausdifferenzierungen und Akzentuierungen – ähnlich wie viele andere anwendungsorientierte Literacy-Diskurse – an mehr oder weniger differenziert begründeten Auflistungen von Skills und Kompetenzen orientieren, die in Anlehnung an Kompetenz-Frameworks oder -Modelle ohne Rekurs auf theoretische Begründungen spezifischer Literacy-Konzeptionen gebündelt werden.

Aus der Perspektive der Medienbildungsforschung sind dabei einige Desiderata auszumachen, wobei sich die analytische Perspektive hier nicht auf ein enges Verständnis von Medienbildungsforschung bezieht, insofern dieses auf konkrete Prozesse, Ergebnisse oder Ziele von KI-bezogener Kompetenzentwicklung ausgerichtet sein kann.

---

6 Die zweite Gruppe wird in der Conclusio irreführender Weise mit dem Label „theoretische Perspektiven“ (Casal-Otero et al. 2023; S. 12) gekennzeichnet, obschon es dabei im Wesentlichen um Implementierungsvorschläge in Bezug auf strategische, curriculare, organisationale und Gender-Aspekte sowie um Fragen der LehrerInnenaus- und -weiterbildung etc. und nicht um theoretische Erwägungen und Klärungen geht.

Auch wenn auf einer solchen Basis durchaus einige Inhalte, mediale Strukturen oder KI-Anwendungsoptionen ergänzt werden können, gründet die analytische Perspektive in einem weiteren Verständnis von Medienbildungsforschung (s. Abb. 3), in dem (meta-)theoretische Überlegungen im Spannungsfeld von Bildungs-, Medien- und Wissenstheorie eine Rolle spielen (vgl. exemplarisch Hug 2007; Bachmair 2010; Marotzki/Meder 2014; Iske et al. 2020; Missomelius 2022).

Einige Desiderata der AI Literacy Diskurse lassen sich wie folgt zusammenfassen:

- *Literacy-Metapher*

Der Ausdruck AI Literacy wird weithin im Sinne einer metaphorischen Anwendung von Konzepten und sozialen Praktiken des Lesens und Schreibens auf den Umgang mit jenen daten- und informationsverarbeitenden Systemen verwendet, die heute gemeinhin mit dem Ausdruck Künstliche Intelligenz verknüpft werden. Die metaphorischen Gehalte und die Modalitäten der figurativen Übertragung der ursprünglichen Bedeutungszusammenhänge werden dabei genauso wenig reflektiert wie die eingangs erwähnten alternativen Basiskonzepte von Literacy und deren Bedeutung im Zusammenhang von KI-bezogener Kompetenzentwicklung.

- *Literacy vs. Numeracy und Mathemacy*

Auch wenn die diskutierten AI Literacy Diskurse auf Lehrkräfte, EntscheidungsträgerInnen und „non-technical learners“ im weiteren Sinne abheben, bleibt aus systematischer Sicht die Frage nach einer Verhältnisbestimmung der Rolle von Literacy, Numeracy und Mathemacy offen. Wenn Literacy nicht im Sinne eines schwach konturierten „All- Inclusive-Pakets“ für unterschiedliche Formen des Wissens und Könnens sowie für variable Bündel von Fähigkeiten und Fertigkeiten verwendet werden soll, braucht es entsprechende Klärungen und Relationierungen.

- *Bildung, Lernen, Wissen und Kompetenz – Literacies und Skills*

In den diskutierten AI Literacy Diskursen spielen bildungstheoretische Ansprüche und Begründungen sowie Bezugnahmen zu den entsprechenden Medienbildungsdiskursen keine Rolle. Die Bündelungen von Skills und Kompetenzen heben durchwegs auf quantitativ-empirisch messbare Outputs ab. Paradoxerweise korrespondiert mit dem zukunftsorientierten qualifikatorischen Anspruch von AI Literacy im Sinne einer materialen Bildung entlang der Trias „Formalisieren – Funktionalisieren – Kontrollieren“ (Stederoth 2024) keine differenzierte Analyse der materiellen Rahmenbedingungen im digitalen Kapitalismus (vgl. Peters 2012; Dander et al. 2021; Eckert/Grünberger 2023). Eine theoretische informierte Verhältnisbestimmung von Bildung, Lernen, Wissen, Kompetenz, Literacy und Skills kommt nicht vor.

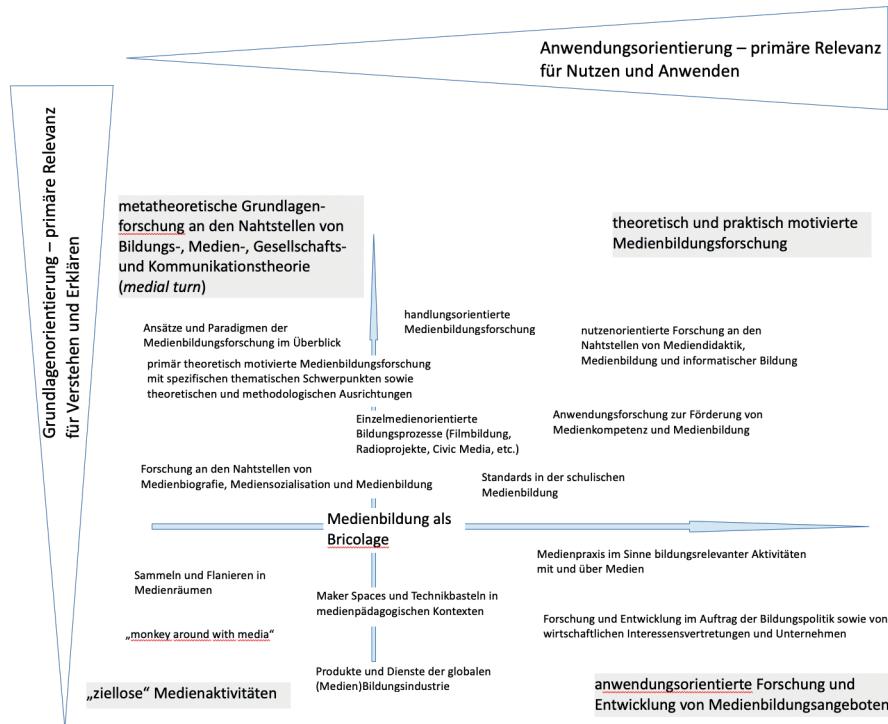


Abbildung 3: Orientierungen der Medienbildungsforschung  
in Anlehnung an Stokes (1997) (eigene Darstellung)<sup>7</sup>

- *Primat der Werkzeugperspektive von KI*

Die favorisierte Werkzeugperspektive von KI korrespondiert in den AI Literacy Diskursen mit einer Unterbewertung historischer, sozio-technischer, gesellschaftlicher und medienkultureller Dimensionen von KI-Entwicklungen.

7 Die Gegenstandsbereiche und Orientierungen der Medienbildungsforschung lassen sich u.a. anhand der Aufteilung wissenschaftlicher Forschungstypen in vier Felder von Donald Stokes (1997) verorten. Auch wenn sein Modell primär im Kontext von forschungs- und förderpolitischen Zusammenhängen eine Aussagekraft hat und methodologische sowie erkenntnis- und wissenstheoretische Dimensionen vernachlässigt werden, bietet die Unterscheidung der vier Quadranten vor allem im Kontext einer Erstbeschäftigung mit der Thematik durchaus einen Orientierungswert. Der „Bohr’sche Quadrant“ (links oben in der Grafik) bezieht sich dabei auf Ansprüche der *Grundlagenforschung* mit differenzierten Erkenntnisinteressen ohne Anwendungsinteressen; „Pasteurs Quadrant“ (rechts oben) bezieht sich auf *anwendungs- bzw. nutzenorientierte Grundlagenforschung*; „Edisons Quadrant“ (rechts unten) meint *angewandte Forschung* ohne epistemologische Ansprüche; der vierte Quadrant (links unten) bleibt bei Stokes (1997) unbestimmt und kann z.B. für unsystematische Erkundungen oder Faktenbestimmungen vorgesehen werden.

Auch kommunikationswissenschaftliche Ausprägungen einer kritischen Medienbildung (vgl. Gapski 2021) sowie Fragen der politischen Ökonomie und der Nachhaltigkeit<sup>8</sup> werden nicht oder nur ungenügend beachtet. Ethische Aspekte werden zwar durchwegs erwähnt, in ihrer Bedeutung allerdings nicht angemessen und vergleichsweise unterkomplex behandelt (vgl. Funk 2022; 2024). Annahmen, dass mit der Beförderung von AI Literacy die vielgestaltigen Problemlösungspotenziale von KI-Anwendungen umfassend zur Geltung gebracht werden können, werden in ihrer Ambivalenz und begrenzten Relevanz nicht angemessen reflektiert. Die Werkzeugperspektive korrespondiert vielmehr mit Vorstellungen der KI-basierten Bearbeitung von Problemen aller Art – vom Lehrkräftemangel bis zur Klima- und Gesundheitskrise –, ohne die begrenzte Relevanz instrumenteller und technologischer Wissensformen ins Kalkül zu ziehen. Fragen der Wissensdiversität und die Anerkennung der Problemlösungsrelevanz unterschiedlicher Wissensformen (vgl. Beinstainer/Dittrich/Hug 2024) zählen klar zu den Desiderata in den AI Literacy Diskursen.

- *Priorität lerntechnologischer Didaktiken*

Die Verweise auf diverse lernpsychologische Konzepte und Lehr-/Lernarrangements suggerieren eine Offenheit in Bezug auf gestaltungsorientierte Didaktiken, expansive Lernformen und die Auswahl von Lern- und Bildungsinhalten, die angesichts der konkreten Ergebnisorientierungen und der rahmenden Digitalisierungsrhetoriken nur sehr begrenzt gegeben ist. Die Selektion der Inhalte und die Priorisierung lerntechnologischer Didaktiken liegt insgesamt im Widerstreit mit Ansprüchen einer Entwicklungsoffenen Zukunftsorientierung. Ansprüche der Zukunftsoffenheit im Bildungswesen setzen priorität auf Freiräume und Ermöglichungsdidaktiken und nicht auf verfügbungsrationalistische Denkweisen und Technologien der Optimierung von Lehr-/Lernprozessen. Die explizite Auseinandersetzung mit altbekannten und neueren Paradoxien in Bildungskontexten kommt in den AI-Literacy-Diskursen zu kurz.<sup>9</sup> Eine (selbst-)kritische Auseinandersetzung mit den gesellschaftlichen Auswirkungen irreführender kybernetischer Ausdrucksweisen (vgl. Krippendorff 2023) fehlt. Ebenso fehlt eine kritische Reflexion diskursiver Anleihen bei der Sozialen Physik in der Tradition von Adolphe Quetelet (1796-1874) (vgl. Adolf/Stehr 2021).

---

8 Vgl. dazu den Beitrag von Oliver Leistert in diesem Band.

9 Vgl. exemplarisch Freiheit und Zwang, Fremd- und Selbstbestimmung, Uniformierung und Vielgestaltigkeit, Mobilisierung und Stabilisierung, Anpassung und Widerständigkeit, Nähe und Distanz, bildungsindustriell vorgefertigte Lernarrangements auf der Basis proprietärer Software und freie Bildungsmedien auf FLOSS-Basis, profitorientierte Strategien der globalen Medienindustrien vs. gemeinwohlorientierte Gestaltung von Bildungsangeboten, etc..

- *Einseitige vs. vielgestaltige Innovationspfade*

Die Relevanzbekundungen hinsichtlich der Förderung von AI Literacy auf breiter Basis gehen Hand in Hand mit einseitigen Vorstellungen eines Innovationspfads der Tech-Giganten. Spielräume für die Kultivierung diverser Innovationspfade im Bildungssystem im Kontrast zur Orientierung an industriellen Einschätzungen der Alternativlosigkeit werden nicht sondiert. Die Rolle von KI und Bildungsrobotik im Dienst verwaltungsbürokratischer und industrieller Interessen muss auch kritisch diskutiert werden, nicht zuletzt angesichts der Ansprüche der Förderung von Kreativität, Zukunftsoffenheit und Innovationsbereitschaft.

- *Medienanthropologische Dimensionen*

Die interdisziplinären Ansprüche der AI-Literacy-Diskurse werden insofern nicht eingelöst als medienanthropologische Dimensionen in ihrer Bedeutung völlig unterschätzt werden. Die historischen Konstellationen interagierender Medien sind für die Bestimmung menschlicher Denk- und Handlungsformen von zentraler und nicht von peripherer Bedeutung (vgl. Margreiter 1999). Insofern KI-Entwicklungen Teil der Medialität der Wahrnehmungs-, Artikulations-, Kommunikations- und Wissensformen sowie der Medienpraktiken geworden sind, ergeben sich neue Dynamiken der Medialisierung und der Mediatisierung, die auch für Bildungsprozesse grundlegend relevant sind. Die Reflexionsbedarfe, die sich im Spannungsfeld qualifikatorischer, humanistischer, anthropozentrischer und post-anthropozentrischer (vgl. Bettinger 2022) Begründungen und Bildungskonzepte auftun, sind auch für die AI Literacy Diskurse von erheblicher Bedeutung.

- *Unbewusste Dimensionen*

Die AI-Literacy-Diskurse haben bewusste und intentionale Prozesse der Förderung von KI-bezogener Kompetenzentwicklung im Blick. Dabei werden diverse Anleihen in der Lern- und Kognitionspsychologie, nicht jedoch in der Sozial- und Tiefenpsychologie gemacht. Unbewusste Prozesse spielen in diesen Diskursen weder in individuellen noch in kollektiven Hinsichten eine Rolle. Fragen nach neuen Über-Ich-Formen, nach Modalitäten der Übertragung und Gegenübertragung, nach neuen Formen der Selbstreflexion in KI-Kontexten oder nach der Adressabilität von Bewusstsein für Kommunikationsprozesse kommen nicht vor, obschon einschlägige Anknüpfungspunkte zu einem vertieften Verständnis der Lern- und Bildungsprozesse beitragen könnten (vgl. Fuchs 1998; Millar 2021; Possati, 2021; Leineweber 2024).

Wie alle Listen, so ließe sich auch diese Liste fortsetzen, ergänzen und weiter ausdifferenzieren. Wichtig ist aufgrund der dargelegten Argumentationen zweierlei: Erstens handelt es sich bei den skizzierten Desiderata nicht um Leerstellen, die mittels Ergänzung von ein paar weiteren Skills oder Kompetenzbereichen leicht bereinigt werden könnten. Eine Neukonzeptionierung von AI Literacy im Sinne einer umfassenden Medienbildung unter Berücksichtigung der KI-Entwicklungen würde ein reflexives Verhältnis zu den Literacy-Diskursen, eine bildungstheoretisch informierte Verhältnisbestimmung sowie eine interdisziplinäre Kooperation der involvierten geistes-, kultur-, sozial- und strukturwissenschaftlichen Fachbereiche erfordern. Der Terminus AI Literacy suggeriert zwar eine Nähe zu bildungs- und kulturwissenschaftlichen Traditionen, die entsprechenden Diskurse sind aber überwiegend technologisch ausgerichtet und werden der Heterogenität der unterschiedlichen Sprachspiele und Wissensformen, die für ein differenziertes Verständnis der komplexen Phänomenbereiche erforderlich sind, nicht gerecht.

Zweitens zeichnen sich die AI-Literacy-Diskurse wie die KI-Diskurse insgesamt durch eine Verwendung von zahlreichen bildsprachlichen Elementen und oft irreführenden Metaphern aus (vgl. Manzeschke & Gransche 2024). Das Spektrum reicht dabei von der Literacy-Metapher bis zur mehr oder weniger selbstverständlichen Zuschreibung menschenähnlicher Züge zu KI-Systemen, Automaten und Robotern (vgl. maschinelles Lernen als Beispiel für einen Anthropomorphismus) und zur Annahme maschinenähnlicher Züge menschlicher Lernprozesse, die in technischen Begriffen beschrieben werden (Robomorphismen). Auch der selbstverständliche Gebrauch von AI im Kompositum AI Literacy ist Teil dieses Problemzusammenhangs, denn „AI is neither *artificial* nor *intelligent*“ (Crawford 2021, S. 8; kursiv i. Org.). Auch wenn die bildsprachlichen Ausdrucksweisen zweifellos eine Brückenfunktion an den Nahtstellen von politischen, wirtschaftlichen, alltagsweltlichen und wissenschaftlichen Diskursen und nicht zuletzt bei der Akquise von Fördermitteln haben, so erübrigt sich damit keineswegs der Bedarf an Aufklärung über normative Aufladungen, begriffspolitische Dimensionen, operative Fiktionen sowie erhellende und verschleiernde Aspekte im Kontext der KI-Rhetoriken.

## Fazit

Das Fazit bleibt insofern ambivalent als die AI-Literacy-Diskurse einen programmatischen Antwortversuch auf Funktions- und Reproduktionserfordernisse gesellschaftlicher Systeme darstellen, der technologische und wirtschaftliche sowie bildungspolitische, organisationale und pädagogische KI-Themen im Sinne eines Zukunftsversprechens in einer breitenwirksamen Weise verknüpft, ohne der Komplexität der Entwicklungen und Herausforderungen angemessen Rechnung zu tragen. Diese Diskurse bieten durchaus relevante Impulse zu den Debatten über „zeitgemäße Bildung“ (Schultz-Pernice 2020), sie werden allerdings dem Anspruch einer umfassenden Medienbildung nicht gerecht. Ihre Nähe zu sozialtechnologischen Varianten der Zukunftsgestaltung liegt im Widerstreit mit Ansprüchen der Zukunftsoffenheit und der Beförderung transformatorischer Potenziale. Inwieweit sich mit AI Literacy trotzdem ein „paradigm case“ sensu Lankshear/Knobel (2007) im Ensemble neuer Literalitäten abzeichnet oder ob wir es eher mit einer Mode und einem Fall der „literacification of everything“ zu tun haben, muss an dieser Stelle offen bleiben.

Der weithin selbstverständliche Gebrauch der unscharfen Begriffe, bildsprachlichen Darstellungsformen und metaphorischen Ausdrucksweisen hinsichtlich „künstlicher Intelligenzen“ unterstreicht den Reflexions- und Aufklärungsbedarf. Für ein differenzierteres Verständnis der Herausforderungen, die mit den daten- und informationsverarbeitenden Systemen verknüpft werden, die heute als „KI-Technologien“ etikettiert werden, bedarf es interdisziplinärer Anstrengungen, die von einer grundsätzlichen Akzeptanz der Problemlösungsrelevanz unterschiedlicher Fachbereiche und Wissensformen ausgehen. Das betrifft nicht zuletzt Fragen der Bildsamkeit von Menschen in individuellen, organisationalen und gesellschaftlichen Hinsichten. Die partielle Aufhebung gesellschaftlicher Bindegänge, die hegemonialen Ansprüche globaler Bildungsindustrien und die Machtverschiebungen im Kontext transversal vernetzter Mediensysteme korrespondieren mit neuen Deutungen von Selbst, Sozial- und Weltverhältnissen in den Kulturen der Digitalität. Die neueren KI-Entwicklungen unterstreichen diesbezügliche Klärungsbedarfe in besonderer Weise. Die AI-Literacy-Diskurse können hier nur begrenzt beitragen, wie die skizzierten Desiderata gezeigt haben. Eine interdisziplinär ausgerichtete und bildungstheoretisch informierte Medienbildungsforschung kann Auswege aus der Falle von KI-Marketing und einseitig verkürzten Denkformen und technofeudalistischen Sackgassen zeigen.

## Literatur

- Adolf, Marian & Stehr, Nico (2021): Information, Wissen und die Wiederkehr der Sozialen Physik. In: Bachor, Martina; Hug, Theo & Pallaver, Günther (Hrsg.): *DataPolitics. Zum Umgang mit Daten im digitalen Zeitalter*. Innsbruck: Innsbruck University Press, S. 36–54. DOI: <https://doi.org/10.25969/mediarep/19959>.
- Almatrafi, Omaima; Johri, Aditya & Lee, Hyuna (2024): A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, 6, 100173. <https://doi.org/10.1016/j.caeo.2024.100173>.
- Bachmair, Ben (Hrsg.) (2010): *Medienbildung in neuen Kulturräumen. Die deutschsprachige und britische Diskussion*. Wiesbaden: VS Verlag.
- Beinsteiner, Andreas; Dittrich, Ann-Kathrin & Hug, Theo (Hrsg.): *Wissensdiversität und formatierte Bildungsräume*. Innsbruck: iup. DOI: <https://doi.org/10.15203/99106-129-8>.
- Bettinger, Patrick (2022): Medienbildung als Transformation relationaler Gefüge. Eine postanthropozentrische Skizze und ihre Tücken. In: Zulaica y Mugica, Miguel & Zehbe, Klaus-Christian (Hrsg.): *Rhetoriken des Digitalen: Adressierungen an die Pädagogik*. Wiesbaden: Springer Vieweg, S. 83-103. [https://doi.org/10.1007/978-3-658-29045-0\\_5](https://doi.org/10.1007/978-3-658-29045-0_5).
- Burgsteiner, Harald; Kandlhofer, Martin & Steinbauer, Gerald (2016): IRobot: Teaching the Basics of Artificial Intelligence in High Schools. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1). <https://doi.org/10.1609/aaai.v30i1.9864>.
- Carolus, Astrid; Koch, Martin J.; Straka, Samantha; Latoschik, Marc Erich & Wienrich, Carolin (2023): MAIIS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1 (2), 100014. <https://doi.org/10.1016/j.chbah.2023.100014>.
- Casal-Otero, Lorena; Catala, Alejandro; Fernández-Morante, Carmen; Taboada, María; Cebreiro; Beatriz & Barro, Senéen (2023): AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*, 10 (29). <https://doi.org/10.1186/s40594-023-00418-7>.
- Crawford, Kate (2021): *Atlas of AI*. New Haven and London: Yale University Press.
- Dander, Valentin, Hug, Theo, Sander, Ina, & Shanks, Rachel (2021): Digital Capitalism, Datafication, and Media Education: Critical Perspectives. *Seminar.Net*, 17(2). <https://doi.org/10.7577/seminar.4493>.
- Eckert, Gregor & Grünberger, Nina (2023): „Bildung und digitaler Kapitalismus“. Hintergründe zur Entstehung und zu einigen zentralen Begriffen. *merz - Zeitschrift für Medienpädagogik*, 04, S. 38–48.
- Ehlers, Ulf-Daniel (2020): *Future Skills: Lernen der Zukunft-Hochschule der Zukunft*. Wiesbaden: Springer.

- Ehlers, Ulf-Daniel; Lindner, Martin & Rauch, Emily (2024): *AIComp. Future Skills für eine von KI beeinflusste Lebens- und Arbeitswelt*. <https://next-education.org/de/projekte/ai-comp/> [Stand vom 31-07-2024].
- EU (2024): *EU Artificial Intelligence Act. Article 4: AI literacy*. Online-Dokument: <https://artificialintelligenceact.eu/article/4/> [Stand vom 30-07-2024].
- Fuchs, Peter (1998): *Das Unbewußte in Psychoanalyse und Systemtheorie. Die Herrschaft der Verlautbarung und die Erreichbarkeit des Bewußtseins*. Frankfurt a.M.: Suhrkamp.
- Funk, Michael (2022): *Roboter- und KI-Ethik. Eine methodische Einführung – Grundlagen der Technikethik*. Band 1. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-34666-9>.
- Funk, Michael (2024): Nichtwissen ist Macht! Künstliche Intelligenz und ihre diversen Ethiken. In Beinsteiner, Andreas; Dittrich, Ann-Kathrin & Hug, Theo (Hrsg.): *Wissensdiversität und formatierte Bildungsräume*. Innsbruck: iup, S. 41-58.
- Gapski, Harald (2021): Künstliche Intelligenz (KI) und kritische Medienbildung. Digitales Deutschland. Online-Dokument. <https://digid.jff.de/kuenstliche-intelligenz-und-kritische-medienbildung-harald-gapski/> [Stand vom 31-07-2024].
- Hug, Theo (2007): Medienpädagogik unter den Auspizien des mediatic turn. In: Sesink, Werner; Kerres, Michael & Moser, Heinz (Hrsg.): *Jahrbuch Medienpädagogik 6. Medienpädagogik – Standortbestimmung einer erziehungswissenschaftlichen Disziplin*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 10-31.
- Hug, Theo (2019): From Literacy to an Ecology of Literacies? Reflections on Some Conceptual Issues. In: Gross, Barbara/Stadler-Altmann, Ulrike (2019): *Beyond erziehungswissenschaftlicher Grenzen. Diskurse zu Entgrenzungen der Disziplin*. Leverkusen: Verlag Barbara Budrich, S. 145-160.
- Hug, Theo (2021): Thesen zum (Un)Behagen in Bildungskulturen der Digitalität. In: Hierdeis, Helmwart (Hrsg.): *Digitalisierung und Gemeinwohl*. Kröning: Asanger, S. 149-171.
- Hug, Theo (2023): Futures Literacy und die Zukunft von Literalitäten. In: Höfer, Hanna; Delere, Malte & Vogel-Lefèbre, Tatjana (Hrsg.): *Bildung, Diversität und Medien. Erkundungen entlang aktueller Diskurslinien. Festschrift für Gudrun Marci-Boehncke*. München: kopaed, S. 109-126.
- Iske, Stefan; Fromme, Johannes; Verständig, Dan & Wilde, Katrin (Hrsg.) (2020): Big Data, Datafizierung und digitale Artefakte. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-28398-8>.
- Krippendorff, Klaus (2023): A critical cybernetics. *Constructivist Foundations*, 19(1): S. 82–93. <https://constructivist.info/19/1/082>.
- Lankshear, Colin & Knobel, Michele (2007): Sampling ‘the new’ in new literacies. In: Knobel, Michele & Lankshear, Colin (Eds.): *A New Literacies Sampler*. New York, NY: Peter Lang, S. 1-24.

- Laupichler, Matthias C.; Aster, Alexandra & Raupach, Tobias (2023): Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education: Artificial Intelligence*, 4. <https://doi.org/10.1016/j.caeari.2023.100126>.
- Leineweber, Christian (2024): Die Transformation des Unbewussten: Selbstwissen unter den Bedingungen der Digitalität. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 63, S. 69-91. <https://doi.org/10.21240/mpaed/63/2024.09.15.X>
- Long, Duri & Magerko, Brian (2020): What is AI Literacy? Competencies and Design Considerations. In Bernhaupt, Regina; Mueller, Florian F.; Verweij, David; Andres, Josh; McGrenere, Joanna; Cockburn, Andy; Avellino, Ignacio; Goguey, Alix; Bjørn, Pernille; Zhao, Shengdong (S.); Samson, Brian P. & Kocielnik, Rafal (Hrsg.): *Proceedings of the 2020 CHI Conference on Human Factors in Computer Systems*. New York, NY: Association for Computing Machinery, S. 1-16. <https://doi.org/10.1145/3313831.3376727>.
- Manzeschke, Arne & Gransche, Bruno (2024): Bilder machen Menschen. Zur Bildermacht der Künstlichen Intelligenz. In: Heinlein, Michael & Huchler, Norbert (Hrsg.): *Künstliche Intelligenz, Mensch und Gesellschaft*. Springer VS, Wiesbaden, S. 109-130. [https://doi.org/10.1007/978-3-658-43521-9\\_5](https://doi.org/10.1007/978-3-658-43521-9_5).
- Margreiter, Reinhard (1999): Realität und Medialität: Zur Philosophie des „Medial Turn“. *Medien Journal*, 23 (1), S. 9-18.
- Marotzki, Winfried & Meder, Norbert (Hrsg.) (2014): *Perspektiven der Medienbildung*. Wiesbaden: Springer.
- Millar, Isabel (2021): *The Psychoanalysis of Artificial Intelligence*. Cham: Springer. <https://doi.org/10.1007/978-3-030-67981-1>.
- Miller, Riel (Ed.) (2018): *Transforming the future: Anticipation in the 21st century*. Paris/Oxford: UNESCO-Routledge.
- Mishra, Punya & Koehler, Matthew J. (2006): Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108, Nr. 6, S. 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>.
- Missomelius, Petra (2022): Bildung – Medien – Mensch. Mündigkeit im Digitalen. Göttingen: V&R unipress.
- Ng, Davy Tsz Kit; Leung, Jac Ka Lok; Chu, Samuel Kai Wah & Qiao, Maggie Shen (2021): Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeari.2021.100041>.
- Ng, Davy Tsz Kit; Xinyu, Chen; Leung, Ka Lok; Chiu, Thomas Kin Fung & Chu, Samuel Kai Wah (2023): Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55 (3), S. 1082-1104. <https://doi.org/10.1111/bjet.13411>.
- Olson, David. R. & Torrance, Nancy (Eds.) (2009): *The Cambridge Handbook of Literacy*. Cambridge: Cambridge University Press.

- Peters, Michael A. (2012): Postmodern Educational Capitalism, Global Information Systems and New Media Networks. *Policy Futures in Education*, 10(1), 23–29. DOI: <https://doi.org/10.2304/pfie.2012.10.1.23>.
- Possati, Luca M. (2021): *The Algorithmic Unconscious. How Psychoanalysis Helps in Understanding AI*. London: Routledge. <https://doi.org/10.4324/9781003141686>.
- Rütti-Joy, Olivia; Winder, Georg & Biedermann, Horst (2023): Building AI Literacy for Sustainable Teacher Education. *Zeitschrift für Hochschulentwicklung*, 18(4), S. 175–189. <https://doi.org/10.21240/zfhe/18-04/10>.
- Schmid, Mirjam & Petko, Dominik (2020): ‹Technological Pedagogical Content Knowledge› als Leitmodell medienpädagogischer Kompetenz. *Zeitschrift MedienPädagogik* 17 (Jahrbuch Medienpädagogik), S. 121–140. <https://doi.org/10.21240/mpaed/jb17/2020.04.28.X>.
- Schultz-Pernice, Florian (2020): Erzählungen für eine Bildung, die an der Zeit ist. Darstellung und Gestaltung von Zeit in narrativen Texten des Anthropozäns. *Medien + Erziehung*, 64 (4), S. 27–35.
- Seldon, Anthony; Metcalf, Timothy & Abidoye, Oladiemji (2020): *The Fourth Education Revolution: Will Artificial Intelligence Enrich or Diminish Humanity?* Buckingham: The University of Buckingham Press.
- Shah, Priten (2023): *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. Hoboken, New Jersey: John Wiley & Sons.
- Stederoth, Dirk (2024): Formalisieren – Funktionalisieren – Kontrollieren. Digitale Bildung als zielgerechte Fortsetzung der Ökonomisierung des Bildungswesens. In: Beinsteiner, Andreas; Dittrich, Ann-Kathrin & Hug, Theo (Hrsg.) (2024): *Wissensdiversität und formulierte Bildungsräume*. Innsbruck: Innsbruck University Press, S. 27–39.
- Stokes, Donald E. (1997): *Pasteur's Quadrant: Basic science and Technological Innovation*. Washington D.C.: Brookings Institution Press.
- Stolpe, Karin & Hallström, Jonas (2024): Artificial intelligence literacy for technology education. *Computers and Education Open*, 6, 100159. <https://doi.org/10.1016/j.caeo.2024.100159>.
- Stordy, Peter (2015): Taxonomy of literacies, *Journal of Documentation*, 71, 3, S. 456–476. <https://doi.org/10.1108/JD-10-2013-0128>.
- Street, Brian V. (1984): *Literacy in theory and practice*. Cambridge: Cambridge University Press.
- Street, Brian V. & Lefstein, Adam (2007): *Literacy: An Advanced Resource Book*. London and New York: Routledge.
- UNESCO (2022): *K-12 AI curricula, A mapping of government-endorsed AI curricula*. Online-Dokument: <https://unesdoc.unesco.org/ark:/48223/pf0000380602> [Stand vom 30-05-2024].

Veith, Hermann (2003): Lernkultur, Kompetenz, Kompetenzentwicklung und Selbstorganisation. Begriffshistorische Untersuchungen zur gesellschaftlichen und pädagogischen Konstruktion von Erziehungswirklichkeiten in Theorie und Praxis. In: Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung e.V./Projekt Qualifikations-Entwicklungs-Management (Hrsg.): *Was kann ich wissen? Theorie und Geschichte von Lernkultur und Kompetenzentwicklung.* (= QUEM-report 82). Berlin, S. 179-229. Abgerufen unter: <http://www.abwf.de/content/main/publik/report/2003/Report-82.pdf> [Stand vom 30-05-2024].

## Kurzbiografien der Mitwirkenden

**Christine Anderl**, Dr. rer. nat.; wissenschaftliche Mitarbeiterin und stellvertretende Leitung der Arbeitsgruppe Alltagsmedien, Leibniz-Institut für Wissensmedien, Tübingen; Forschungsinteressen: Vernetzung und soziale Interaktion über digitale Medien, Einsatz von KI-Algorithmen im beruflichen und privaten Alltag.

**Nico Ehrhardt**, MSc; wissenschaftlicher Mitarbeiter und Doktorand in der Arbeitsgruppe Alltagsmedien am Leibniz-Institut für Wissensmedien, Tübingen; Arbeitsschwerpunkte: KI-Wahrnehmung, Diskriminierung durch KIs.

**Michael Funk**, Univ.-Ass. (Post Doc) Dr. phil. B.A. M.A.; RG Cooperative Systems, Fakultät für Informatik, Universität Wien; Technikethik, KI-Ethik, Philosophie der Informatik, Methodik der Transdisziplinarität.

**Irene Häntscher-Erhart**, Mag. Dr., Vizerektorin für Digitalisierung und Nachhaltigkeit, Universität Innsbruck, studierte Wirtschaftsinformatik an der Johannes Kepler Universität Linz. Dort sowie an der Universität St. Gallen war sie mehrere Jahre als wissenschaftliche Projektleitung tätig, bevor sie in die Privatwirtschaft wechselte. Nach mehr als zwei Jahrzehnten und verschiedenen (Führungs-)Positionen sowie externer Lehrtätigkeit an verschiedenen Institutionen kehrte sie 2019 an die Fachhochschule Vorarlberg und somit in den Hochschulsektor zurück. Sie leitete die den Bachelorstudiengang Informatik – Digital Innovation und den Masterstudiengang Wirtschaftsinformatik – Digital Transformation sowie das Kompetenzfeld Data and Processes. Mit März 2023 übernahm sie das Amt der Vizerektorin für Digitalisierung und Nachhaltigkeit an der Universität Innsbruck.

**Theo Hug**, Dr. phil., Professor für Erziehungswissenschaft am Institut für Medien, Gesellschaft und Kommunikation der Universität Innsbruck mit Schwerpunkt Medienpädagogik und Kommunikationskultur. Gemeinsam mit Josef Mitterer ist er Nachlassverwalter des Ernst von Glasersfeld Archivs. Zu seinen Arbeitsschwerpunkten zählen Medienpädagogik, Medienbildung, mobiles Lernen, Bildungs- und Wissenstheorie sowie Methodologie und Wissenschaftsphilosophie.

**Stefanie H. Klein**, MA; wissenschaftliche Mitarbeiterin und Doktorandin in der Arbeitsgruppe Alltagsmedien am Leibniz-Institut für Wissensmedien, Tübingen; Arbeitsschwerpunkte: Akzeptanz von Chatbots, Mensch-Maschine-Kommunikation.

**Oliver Leistert**, Dr. phil. Mag. phil., Universitätsassistent am Institut für Medien, Gesellschaft und Kommunikation der Universität Innsbruck; gegenwärtige Forschungs- und Arbeitsinteressen: Geschichte und Theorie digitaler Medien, Medien- und Techniktheorien, Social Media und ihre Folgen für Subjekt und Gesellschaft, Kryptographie aus kultur- und sozialwissenschaftlicher Sicht, Algorithmuskulturen, Surveillance Studies, Sustainable Computing.

**Eike J. Meyer**, Dr., forscht zu generativer KI in Text-, Bild und Videogeneratoren, betrieblichen Anwendungsfällen und AI Leadership am Forschungszentrum der Fachhochschule Kiel und leitet dort im KI Anwendungszentrum die Fachgruppe Generative KI. Er war über 12 Jahre bei IBM als IT Berater und Manager aktiv, mit dem Schwerpunkt auf KI-Anwendungen im betrieblichen Umfeld. Daneben hat er als Geschäftsführer und Gründer des Brettspielverlags Fyrnwest Games die Adaption generativer KI-Werkzeuge in der Brettspielindustrie vorangetrieben und adaptiert auch als Lehrbeauftragter regelmäßig den Einsatz von KI-Tools in der Lehre.

**Petra Missomelius**, Dr. phil., Professorin am Institut für Medien, Gesellschaft und Kommunikation an der Universität Innsbruck. Arbeits- und Forschungsschwerpunkte: Medientheorie, (Bewegt-)Bildkulturen und Medienpraktiken, Körper- und Techniktheorien, Medien und Wissenschaft/Bildung. Publikation „Bildung-Medien–Mensch: Mündigkeit im Digitalen“ (2022). Sprecherin des Forschungszentrums *Innsbruck Media Studies*.

**Heike Ortner**, Dr. phil., ist assoziierte Professorin am Institut für Germanistik an der Universität Innsbruck. Ihre Forschungsschwerpunkte sind multimodale Diskurse im Kontext von Identitäts- und Wissenskonstruktion, Sprache und Emotion, interpersonale und institutionelle Gesundheitskommunikation. 2023 Publikation der Habilitation „Sprache – Bewegung – Instruktion. Multimodales Anleiten in Texten, audiovisuellen Medien und direkter Interaktion“ (De Gruyter).

**Hermann Petz**, Mag., Jahrgang 1961, ist seit 2003 Vorstandsvorsitzender des Tiroler Medienhauses Moser Holding AG sowie Mitglied des Vorstandes und des Präsidiums des Verbands Österreichischer Zeitungen (VÖZ), Vorstandsvorsitzender der Austria Presse Agentur (APA) und Österreich-Delegierter der ENPA (European Newspaper Publishers Association).

**Justus Piater**, Prof., Ph.D., heads the Digital Science Center of the University of Innsbruck and the Intelligent and Interactive Systems research group at the Department of Computer Science. The goal of his research is to enable autonomous robots to perceive and act flexibly and robustly in unstructured environments, leveraging machine learning methods to build perceptual, motor and reasoning skills.

**Clemens Pig**, Dr., ist Vorsitzender der Geschäftsführung und geschäftsführender Vorstand der APA – Austria Presse Agentur Unternehmensgruppe (Wien) und Vize-Präsident des Verwaltungsrates der Keystone-SDA-ATS AG (Bern) sowie Präsident der Vereinigung der unabhängigen Nachrichtenagenturen Europas (EANA).

**Maya Pindeus**, Mag.Arch, MSc, MA, KI Unternehmerin, Architektin und Designerin mit Fokus auf die Interaktion von Mensch und Maschine. Sie ist Mitgründerin der KI-Unternehmen Humanising Autonomy und Another Earth, die ethische Künstliche Intelligenz entwickeln, die sich positiv auf Mensch und Gesellschaft auswirken soll. Maya ist Absolventin der Universität für Angewandte Kunst in Wien, dem Imperial College London und dem Royal College of Art in London. Sie ist "Forbes 30 under 30" - Preisträgerin und ihre Arbeit in der KI- und Technologiebranche hat weltweit Anerkennung gefunden. Arbeitsschwerpunkte: KI- Implementierung, Ethische - und Verantwortungsbewusste KI, Synthetische Daten für KI-Training.

**Miro Roman**, Assistant Professor at the House of Coded Objects at studio2 at the University of Innsbruck, and a senior lecturer and researcher at Studio Meteora at the chair for Digital Architectonics at ETH Zurich. He was a guest lecturer at ATTP, TU Wien, and a researcher at the Future Cities Laboratory, the interdisciplinary research program of the Singapore ETH Centre, where he co-edited A Quantum City book. Miro conceptualized and developed the computational library Xenotheke and its search instrument "Ask Alice". Alice\_ch3n81 and Miro together wrote "A Play Among Books".

**Sonja Utz**, Prof. Dr.; Leitung der Arbeitsgruppe Alltagsmedien, Leibniz-Institut für Wissensmedien & Professorin für Kommunikation mittels sozialer Medien an der Universität Tübingen; Arbeitsschwerpunkte: wissensbezogene Nutzung sozialer Medien, Mensch-Maschine-Kommunikation, Algorithmenakzeptanz.

**Doris Weßels**, Prof. Dr., Professorin für Wirtschaftsinformatik mit den Schwerpunkten Projektmanagement und Natural Language Processing an der Fachhochschule Kiel. Seit 2021 ist sie Mitglied im KI-ExpertLab Hochschullehre des deutschen KI-Campus (<https://ki-campus.org/blog/ki-expertlab-hochschullehre>) und hat dort die Themengruppe „KI und Academic Writing“ initiiert. Zum 1.09.2022 hat sie gemeinsam mit dem assoziierten Partner KI-Campus und weiteren Expert\*innen und Lehrenden deutscher Hochschulen aus den Bereichen Informatik, Bildungsforschung und Schreibdidaktik das Virtuelle Kompetenzzentrum „Künstliche Intelligenz und wissenschaftliches Arbeiten - Schreiben lehren und lernen mit KI“ (<https://www.vkkiwa.de/>) gegründet. Sie ist Mitglied in diversen KI-Experten- und Beiratsgremien, u.a. der Stadt Wien, und berät Ministerien und Bildungsinstitutionen bei der Förderung von Digitalkompetenzen.

Weitere Informationen: [https://de.wikipedia.org/wiki/Doris\\_We%C3%9Fels](https://de.wikipedia.org/wiki/Doris_We%C3%9Fels)

Die Popularität von Tools wie bHuman, Dall-E2 und ChatGPT demonstriert, wie schnell KI-Anwendungen zu einem selbstverständlichen Teil unserer Gesellschaft und unserer Umwelt werden. Sind sie als neue Werkzeuge zu verstehen, die uns helfen, der zunehmenden Komplexität unserer Gesellschaft zu begegnen? Oder verändert der unmerkliche Einsatz dieser Technologien auch uns selbst, unser Menschen- und unser Selbstbild?

Einsatzmöglichkeiten und Fähigkeiten der KI-Tools sind bereits jetzt erstaunlich und es ist noch kein Ende in Sicht. Möglicherweise helfen sie uns dabei, die Probleme unserer Zeit zu lösen und nachhaltigere Gesellschafts- und Wirtschaftsformen zu entwickeln. Doch ertönt aus diesen ökonomischen und ökologischen Zusammenhängen auch die Forderung nach politischen Regularien. Gibt es Grenzen der Verantwortung und der Ethik? Müssen wir regulieren, was KI darf und was nicht? Und wie können wir das gewährleisten? Die Beiträge setzen sich mit diesen Fragen sowie mit den Möglichkeiten und Herausforderungen der Nutzung künstlicher Intelligenz im Medienalltag und darüber hinaus auseinander und beleuchten Themen, die unsere Gesellschaft heute, aber vor allem auch in Zukunft beschäftigen werden.

ISBN 978-3-99106-139-7



9 783991 061397