

Tobias Holischka
Carolin Kreisbeck
Christina Pfeuffer
Michael Winklmann (Hg.)

Künstliche Intelligenz

Chancen, Herausforderungen,
Perspektiven

wbg Academic

HERDER

Tobias Holischka / Carolin Kreisbeck / Christina Pfeuffer / Michael Winklmann (Hg.)

Künstliche Intelligenz

KU University Press
Band 7

Tobias Holischka / Carolin Kreisbeck / Christina Pfeuffer /
Michael Winklmann (Hg.)

Künstliche Intelligenz

Chancen, Herausforderungen,
Perspektiven

wbgAcademic

HERDER 

FREIBURG · BASEL · WIEN

Parallele Veröffentlichung auf dem Repository der Katholischen Universität Eichstätt-Ingolstadt (KU.edoc):
<https://doi.org/10.17904/ku.edoc.36387>

wbg Academic ist ein Imprint der Verlag Herder GmbH
© Verlag Herder GmbH, Freiburg im Breisgau 2026
Hermann-Herder-Straße 4, 79104 Freiburg
Alle Rechte vorbehalten
www.herder.de

Bei Fragen zur Produktsicherheit wenden Sie sich an:
produktsicherheit@herder.de

Satz: Michael Winklmann
Umschlaggestaltung: Arnold & Domnick GbR, Leipzig
Umschlagmotiv: © Beckett LeClair / <https://betterimagesofai.org> / <https://creativecommons.org/licenses/by/4.0/>

Printed in Germany

ISBN Print: 978-3-534-64351-6
ISBN E-Book (OA): 978-3-534-64352-3

Dieses Werk ist mit Ausnahme der Abbildungen (Buchinhalt und Umschlag) als Open-Access-Publikation im Sinne der Creative-Commons-Lizenz CC BY-NC-ND International 4.0 („Attribution-NonCommercial-NoDerivatives International“) veröffentlicht. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie <http://creativecommons.org/licenses/by-nc-nd/4.0/>. Jede Verwertung in anderen als den durch diese Lizenz zugelassenen Fällen bedarf der vorherigen schriftlichen Einwilligung des Verlages.

Inhalt

Vorwort.....	7
Maschinen, die lernen: Ein Blick in die Welt des Machine Learning.....	9
Interview mit Felix Voigtlaender	
ChatGPT, der „Tod des Autors“ und die Geltung von Texten	17
Lukas Ohly	
I create paths that lead to the clouds in which we go: Die KI als Partnerin in der zeitgenössischen Medienkunst	35
Ornella Fieres	
KI und die vierte industrielle Revolution	49
Interview mit Christoph Kreisbeck	
„Nun sag, wie hast du’s mit KI?“: Wissenschaftstheoretische und kartographische Orientierungen zur religionspädagogischen Forschung über Künstliche Intelligenz.....	57
Mariusz Chrostowski/Johannes Heger	
Die Struktur des Digitalen: Lernen über KI an Lernstationen	81
Michael Köck	
Verzeichnis der AutorInnen und InterviewpartnerInnen	103

Vorwort

Im Wintersemester 2022/23 veranstaltete die Katholische Universität Eichstätt-Ingolstadt eine interdisziplinäre Ringvorlesung mit dem Titel „Are you scared yet, human? Erwartungen, Realitäten und Perspektiven Künstlicher Intelligenz“. Die Veranstaltungsreihe sollte als Plattform dienen, um Perspektiven aus unterschiedlichen Disziplinen – von Philosophie und Theologie über Mathematik und Informatik bis hin zu Psychologie und Soziologie – zu vernetzen. Ziel war es, sowohl die technischen Möglichkeiten als auch die ethischen, gesellschaftlichen und anthropologischen Implikationen der KI-Entwicklung zu beleuchten.

Kurz nach Beginn der Vorlesungsreihe erhöhte sich die Aktualität des Themas noch einmal: Mit der Veröffentlichung von ChatGPT im November 2022 entfachte sich ein medialer und gesellschaftlicher Hype um Künstliche Intelligenz, der bis heute ungebrochen ist.

Der gewählte Titel der Ringvorlesung greift einen Satz aus einem Text auf, der am 8. September 2020 im britischen Guardian erschien – vollständig von einer KI verfasst. Der Text sorgte in Fachkreisen und der Öffentlichkeit für intensive Debatten: Wie autonom können Algorithmen wirklich schreiben? Welche Rolle spielt menschliche Kreativität noch?

Fünf Jahre später gibt es eine Fülle von Tools, die KI-gestützte Texterstellung für Alltags- und Bildungszwecke anbieten – von der Erstellung von Gebrauchstexten bis hin zu Referaten in Schulen und Seminararbeiten an Universitäten. Die Grenzen zwischen menschlicher und maschineller Intelligenz verschwimmen zunehmend, und nicht nur die Hochschulbildung steht vor der Herausforderung, mit diesen Entwicklungen Schritt zu halten.

Die rasante technische Entwicklung im Bereich der Künstlichen Intelligenz erfordert nicht nur technisches Verständnis, sondern auch eine gründliche Reflexion über ihre Folgen. Dieser Band möchte diesem Spannungsfeld gerecht werden: Die Beiträge zeigen einerseits die bahnbrechenden Fortschritte auf dem Gebiet auf. Auf der anderen Seite werden auch grundlegende Fragen nach der Natur menschlicher Intelligenz, nach ethischen Grenzen der KI-Nutzung und nach der Verantwortung, die mit solchen Technologien einhergeht, gestellt.

Der Band vereint nicht nur klassische wissenschaftliche Aufsätze, sondern auch Interviews mit ExpertInnen, die von Maximilian Weidmann geführt wurden. Diese Interviews bieten einen direkten Zugang zu den Stimmen von ForscherInnen, die an der Schnittstelle von Technologie und Gesellschaft arbeiten. Sie ergänzen die theoretischen Beiträge um praktische Einsichten.

Wir danken dem wissenschaftlichen Beirat der KU-University Press für die Aufnahme in die Reihe und wbg Academic im Verlag Herder für die Betreuung der Publikation. Der Hochschulleitung der Katholischen Universität Eichstätt-Ingolstadt sind wir für die Übernahme des Druckkostenzuschusses sehr dankbar.

Eichstätt, im Herbst 2025

Tobias Holischka, Carolin Kreisbeck, Christina Pfeuffer und Michael Winkmann

Maschinen, die lernen

Ein Blick in die Welt des Machine Learning

Interview mit Felix Voigtländer

KI, also Künstliche Intelligenz, begegnet uns zu einem gewissen Grad überall im Alltag. Sie beschäftigen sich in ihrem Fachbereich mit Machine Learning. Was ist Machine Learning? Ist das das gleiche wie künstliche Intelligenz?

Das Grundproblem der Künstlichen Intelligenz liegt darin, dass es sich um einen nicht besonders präzise definierten Begriff handelt, der im Wesentlichen bedeutet, einen Computer dazu zu bringen, Aufgaben auszuführen, für die normalerweise menschliche Intelligenz erforderlich ist. Die Frage, ob beispielsweise ein Taschenrechner als Künstliche Intelligenz einzustufen ist, bleibt dabei jedoch unklar. Künstliche Intelligenz stellt den umfassenden Oberbegriff dar, wobei Maschinelles Lernen ein Teil davon und Deep Learning wiederum ein spezifischer Teilbereich innerhalb des Maschinellen Lernens ist. Meiner Meinung nach hat sich der Begriff Deep Learning in letzter Zeit ein wenig verschoben. Wenn in den Medien von Künstlicher Intelligenz gesprochen wird, bezieht sich dies häufig auf den Bereich des Deep Learning, der in jüngster Zeit für bedeutende, publikumswirksame Durchbrüche verantwortlich war.

Zentraler Begriff im Bereich der KI ist der Algorithmus. Was ist eigentlich ein Algorithmus?

Ein Algorithmus ist eine sehr präzise und detaillierte Anweisungsfolge, die, wenn genau befolgt, zu einem bestimmten Ergebnis führt. Ein einfaches Beispiel dafür wäre das Kaffeekochen mit einer Kaffeemaschine: Zuerst füllt man Wasser ein, dann steckt man die Kaffeemaschine ein, stellt den Becher darunter und schaltet sie schließlich ein. Dieser Vorgang kann als ein umgangssprachlich formulierter Algorithmus betrachtet werden. Im Allgemeinen bezieht sich der Begriff Algorithmus jedoch auf computerimplementierte Anweisungen. Ein Algorithmus kann nur sehr spezifische, kleine Schritte wie „Überprüfe, ob X gleich drei ist“ oder „Erhöhe X um 42“ ausführen. Die Informatik beschäftigt sich überwiegend mit der Entwicklung effizienter Algorithmen, um bestimmte Aufgaben wie das Sortieren einer Liste oder das Finden des kleinsten Elements einer Liste zu lösen.

Ist dann ein Algorithmus, einfach gesagt ein Computerprogramm?

Es ist vielleicht ein bisschen abstrakter. Ein Algorithmus ist im Grunde eine allgemeine Folge von Anweisungen, die eine bestimmte Aufgabe löst. Er kann als eine Liste von Schritten betrachtet werden, die in einer bestimmten Reihenfolge ausgeführt werden müssen, um ein bestimmtes Ergebnis zu erzielen. Die Besonderheit eines Algorithmus liegt darin, dass er unabhängig von der konkreten Implementierung ist. Das bedeutet, dass derselbe Algorithmus in verschiedenen Programmiersprachen wie Python oder C++ umgesetzt werden kann, aber

die zugrunde liegende Logik und der Ablauf der Schritte bleiben dieselben. Der Algorithmus selbst bleibt also unverändert, während die konkrete Implementierung je nach Programmiersprache und Umgebung variieren kann.

Sprechen wir da von Zahlen Reihungen mit Null und Eins oder von Worten?

Ein Algorithmus ist eine Beschreibung in Worten, die als Pseudo-Code bezeichnet wird. Dieser Pseudo-Code beschreibt die Schritte und Logik des Algorithmus in einer verständlichen und lesbareren Form. Wenn dieser Algorithmus dann in eine Programmiersprache übersetzt wird, entstehen ebenfalls Zeichenketten oder Worte, die zwar maschinenlesbar sind, aber immer noch eine gewisse Verständlichkeit für den Menschen aufweisen.

Letztendlich kann alles, was auf einem Computer gespeichert oder verarbeitet wird, einschließlich Bilder, Filmen und Texten, auf eine Folge von Nullen und Einsen reduziert werden. Das ist die grundlegende Repräsentationsform für digitale Informationen. Allerdings denkt man normalerweise nicht auf dieser Ebene über Algorithmen nach.

Wie kann man sich einen Algorithmus an einem kurzen Beispiel vorstellen?

Ein einfaches Beispiel dafür wäre folgendes: Wenn man eine Liste von Zahlen hat und die Größte bestimmen möchte, denkt man als Mensch zunächst: Wenn ich zehn Zahlen habe, kann ich direkt sehen, welche die Größte ist. Bei einer Liste mit 1 Million Einträgen ist dies jedoch nicht mehr so einfach, auch nicht für einen Computer. Um dies zu lösen, geht man die Liste Schritt für Schritt durch. Man beginnt mit dem ersten Element, das zunächst als das Größte gilt, das man bisher gesehen hat, und schreibt es sich auf. Dann geht man zum nächsten Element, indem man den Index um eins erhöht, und überprüft, ob es größer ist als das bisher größte Element, das man sich gemerkt hat. Wenn ja, ersetzt man das größte Element durch das neue. Wenn nicht, geht man zum nächsten Element. Dieser Prozess wird fortgesetzt, bis man das Ende der Liste erreicht hat. Am Ende hat man sich dann das größte Element gemerkt.

Wenn Sie diesen Begriff Machine Learning kurz einem Fachfremden beschreiben müssten, wie würden Sie ihn definieren?

Um ein Problem mittels Machine Learning zu lösen, beginnt man damit, eine große Menge an Beispieldaten („Trainingsdaten“) zu sammeln. Machine Learning besteht darin, dass der Computer auf der Grundlage dieser Daten lernt, das eigentliche Problem möglichst gut zu lösen. Beispiele hierfür sind die Bilderkennung, die Transkription von Text aus Tonaufnahmen oder die Erkennung von Spam.

Gibt der Mensch bei Machine Learning die Befehle einem Programm oder lernt das Programm selbstständig?

Sowohl als auch. Der Mensch sammelt die Trainingsdaten und wählt den geeigneten Machine Learning-Algorithmus aus. Anschließend übernimmt die Maschine und lernt mithilfe

des ausgewählten Algorithmus und der eingegebenen Daten, um hoffentlich die richtige Lösung zu finden.

Wie unterscheidet sich denn das Lernen beim Menschen von dem Lernen einer Maschine?

Wenn man von Lernen spricht, muss man zwischen verschiedenen Arten des Lernens unterscheiden. Insbesondere an der Universität assoziiert man Lernen oft mit Auswendiglernen, wie zum Beispiel das Memorieren von Vokabeln oder Fakten. Dies ist jedoch nicht das, was im Kontext von Machine Learning gemeint ist. Ein Computer kann problemlos große Mengen an Daten speichern, aber das Ziel von Machine Learning ist es, durch Beispiele zu lernen, wie man bestimmte Aufgaben ausführt. Ich bin kein Experte für das menschliche Lernen. Aber viele Aspekte von Machine Learning, insbesondere Deep Learning, sind von den Funktionsweisen des menschlichen Gehirns und neuronalen Netzen inspiriert. Einen wesentlichen Unterschied gibt es aber: Menschen verfügen über umfangreiches implizites Vorwissen über die Welt, wie zum Beispiel die Orientierung in Raum und Zeit oder das Verständnis von Kausalitäten. Im Gegensatz dazu verfügt ein Computer oder ein Machine Learning-Algorithmus über keinerlei Vorwissen. Er erhält lediglich die Daten, die ihm zur Verfügung gestellt werden, und muss auf Grundlage dieser Daten lernen – ohne irgendeine Vorstellung von der Welt zu haben. Ein Beispiel dafür wäre das Training eines Algorithmus mit einer Menge von E-Mails, die jeweils als Spam oder nicht-Spam gekennzeichnet sind. Der Algorithmus muss dann allein auf Basis dieser Daten lernen, um Spam-E-Mails zu erkennen, ohne jegliches Vorwissen über die Welt oder den Kontext, in dem diese E-Mails verschickt wurden.

Es gibt zwei grundlegende Ansätze, um einem Menschen etwas beizubringen. Der erste Ansatz besteht darin, einen Algorithmus oder eine schrittweise Anleitung zu vermitteln, wie zum Beispiel die Multiplikation von Zahlen. Hierbei wird die Funktionsweise und die Reihenfolge der Schritte erklärt. Der zweite Ansatz besteht darin, Beispiele zu zeigen, um das Konzept oder die Unterscheidung zu vermitteln. Ein Beispiel dafür wäre, wenn man jemandem drei Bilder von Donald Trump und drei Bilder von Joe Biden zeigt, um ihm zu helfen, die beiden Personen zu unterscheiden. Nachdem man diese Beispiele gesehen hat, kann man ein neues Bild erkennen und die entsprechende Person identifizieren. Dieser zweite Ansatz ist genau das, was die Machine Learning-Revolution bei Computern ermöglicht hat. Früher musste man Computern explizite Anweisungen geben, wie zum Beispiel „um das kleinste Element in einer Liste zu finden, führe diese und diese Schritte aus“. Mit Machine Learning kann man dem Computer jedoch anhand von Beispielen etwas beibringen, ähnlich wie man es bei Menschen tut. Allerdings gibt es einen wesentlichen Unterschied: Während es für einen Menschen oft ausreicht, nur wenige Beispiele zu sehen, um ein Konzept zu verstehen, benötigt ein Computer in der Regel tausende bis Millionen von Beispielen, um ein bestimmtes Muster oder eine bestimmte Aufgabe zu erlernen. Dies liegt daran, dass der Computer kein Vorwissen oder keine impliziten Annahmen über die Welt hat und daher auf eine viel größere Menge an Daten angewiesen ist, um zu lernen.

An welche Grenzen stößt denn dann Maschine Learning?

Das Machine Learning funktioniert in den letzten 5 bis 10 Jahren so gut, weil wir jetzt Zugang zu viel größeren Datenmengen haben als früher und auch über viel mehr Rechenleistung verfügen. Diese beiden Faktoren waren früher die großen limitierenden Faktoren für die Entwicklung von Machine Learning-Modellen. Tatsächlich haben Forscher auch früher mit neuronalen Netzen gearbeitet, was jetzt wieder sehr populär ist. Allerdings funktionierten diese Modelle damals nicht sehr gut, weil es einfach nicht genug Daten und Rechenleistung gab, um sie effektiv zu trainieren. Heute gibt es Modelle wie ChatGPT, die sehr beeindruckende Ergebnisse erzielen, aber auch sehr teuer sind, um sie zu betreiben. Die Kosten für den Betrieb solcher Modelle können in die Zehntausende von Dollar pro Tag reichen, einfach weil sie so viel Rechenleistung benötigen, um Vorhersagen zu treffen. Ein weiteres Problem bei neuronalen Netzen ist, dass es sehr schwierig ist, zu verstehen, was sie intern machen. Sie machen andere Fehler als Menschen und können oft unerwartete Entscheidungen treffen. Ein Beispiel dafür ist ein selbstfahrendes Auto, das einen blauen LKW wegen seiner blauen Farbe für den Himmel hält und deswegen in ihn hineinfährt. Ein solcher Fehler wäre für einen Menschen unmöglich, weil unsere Intelligenz auf anderen Prinzipien basiert. Die Fehler, die von neuronalen Netzen gemacht werden, sind oft unerwartet und können schwer zu erklären sein.

Kann Machine Learning salopp gesagt dann alle menschlichen Limitierungen überwinden?

Ein großer Vorteil von Machine Learning-Algorithmen ist ihre Fähigkeit, natürlicherweise viel größere Datenmengen zu verarbeiten als ein einzelner Mensch und darin liegende Muster zu erkennen und auszunutzen, die für Menschen möglicherweise nicht erkennbar sind. Zum Beispiel macht Wikipedia wahrscheinlich nur etwa 3 % der gesamten Trainingsdaten für Chat-GPT aus, was die gigantische Größe der Datenmenge verdeutlicht. Ähnlich verhält es sich mit Programmen, die das Spielen von Go erlernt haben: Innerhalb von fünf Stunden konnten sie mehr Spiele spielen, als ein Mensch in seinem ganzen Leben spielen könnte. Eine Grenze dieser Algorithmen ist jedoch, dass sie keine wirkliche Intelligenz besitzen. Stattdessen lernen sie, anhand der gegebenen Daten Regelmäßigkeiten zu erkennen und diese auszunutzen. Dies kann manchmal den Eindruck erwecken, als seien sie intelligent, was jedoch nicht der Fall ist. Es ist wichtig, sich dies immer vor Augen zu halten: Eine Maschine ist nicht unfehlbar. Wenn man einen Taschenrechner verwendet, rechnet dieser korrekt, zum Beispiel 7 mal 8 ergibt 56. Dies unterscheidet sich jedoch grundlegend von den Machine Learning-Algorithmen, die auf der Grundlage von Daten lernen und daher auch falsche oder seltsame Entscheidungen treffen können, wenn die Daten fehlerhaft sind. Dies ist ein wichtiger Unterschied zu dem, was man normalerweise von Computern gewohnt ist, nämlich dass sie immer das Richtige tun.

In ihrem Vortrag sprechen Sie auch von Supervised Machine Learning? Wer überwacht da wen?

In diesem Fall überwacht der Mensch den Computer oder den Machine Learning Algorithmus. Im Fall des Supervised Machine Learning ist ein Beispiel die Spam-Erkennung. Dabei

gibt man dem Computer eine große Menge an E-Mails und klassifiziert jede einzelne als „Spam“ oder „nicht Spam“. Der Begriff „Supervised“ bedeutet genau, dass man dem Computer für die Trainingsdaten die richtigen Antworten vorgibt, damit er aus diesen Beispielen lernen kann. Im Gegensatz dazu gibt es auch unsupervised Methoden, die weniger verbreitet sind. Bei dieser Methode gibt man dem Computer nur die E-Mails, ohne sie als „Spam“ oder „nicht Spam“ zu kennzeichnen. Der Computer kann dann selbstständig Cluster, also Gruppen von ähnlichen E-Mails, finden. Anstatt nach Spam oder Nicht-Spam zu suchen, könnte der Computer beispielsweise private und Arbeitsmails voneinander unterscheiden. Der wesentliche Unterschied zwischen diesen beiden Ansätzen liegt darin, dass bei supervised Methoden die korrekten Antworten vorgegeben werden, während bei unsupervised Methoden der Computer selbst Muster finden soll, ohne dass man ihm vorher sagt, wonach er genau suchen soll.

Wenn dann doch mal eine Mail durchrutscht, die doch Spam ist und nicht gefiltert wird, müssen dann neue Trainingsdaten hinzugegeben werden, damit so was dann verbessert werden kann?

Genau, das ist ein weiterer Vorteil von einigen Machine Learning-Algorithmen, besonders solchen die auf neuronalen Netzen basieren. Sobald sie einmal entwickelt und trainiert sind, können sie prinzipiell weiterhin lernen und sich verbessern. Jedes neue Trainingsdatum, wie zum Beispiel eine neu markierte Spam-E-Mail, kann verwendet werden, um den Algorithmus zu aktualisieren und seine Genauigkeit zu erhöhen. Wenn Sie also eine E-Mail in Ihrem E-Mail-Ordner erhalten und sie als Spam markieren, kann der Algorithmus aus dieser Erfahrung lernen und nächstes Mal ähnliche E-Mails mit höherer Wahrscheinlichkeit als Spam identifizieren. Die Entwicklung von Spamfiltern ist ein interessantes Beispiel dafür, wie sich die Technologie im Laufe der Zeit weiterentwickelt hat. Früher basierten Spamfilter auf einfachen Regeln, die bestimmte Wörter oder Phrasen mit einer bestimmten Wahrscheinlichkeit für Spam oder Nicht-Spam assoziierten.

Zu Beginn ging es auch um Deep Learning. Was ist Deep Learning?

Deep Learning ist ein Teilbereich von Machine Learning, der sich auf eine spezielle Art von Algorithmen konzentriert, nämlich künstliche neuronale Netze, insbesondere tiefe künstliche neuronale Netze. Ein neuronales Netz besteht aus vielen Neuronen, die jeweils eine bestimmte Berechnung durchführen. Wenn ein neuronales Netz tief ist, bedeutet das, dass es viele Schichten von Neuronen gibt, die nacheinander arbeiten. Die Eingabe bildet die erste Schicht, die dann eine bestimmte Berechnung durchführt und eine Ausgabe produziert. Die nächste Schicht nimmt diese Ausgabe als Eingabe und führt eine weitere Berechnung durch, um eine neue Ausgabe zu produzieren. Dieser Prozess wiederholt sich viele Schichten lang, wobei jede Schicht die Ausgabe der vorherigen Schicht als Eingabe verwendet.

Ist das dann ein Netz von Maschinen?

Nein. Es läuft alles im Computer ab, ohne dass eine physikalische Maschine benötigt wird. Der Computer simuliert ein künstliches neuronales Netz. Ein künstliches Neuron kann man

sich wie folgt vorstellen: Es nimmt bestimmte Eingaben entgegen. Ein einfaches Beispiel wäre ein künstliches Neuron, das vorhersagt, ob ein Student eine Klausur bestehen wird oder nicht. Die Eingaben könnten dabei die Anzahl der Lernstunden und die Abiturnote sein. Als Ergebnis kommt dann eine Ausgabe heraus. Ein einzelnes Neuron ist jedoch viel zu einfach, um damit komplexere Sachverhalte darstellen zu können. Künstliche neuronale Netze bestehen aus Millionen oder sogar Milliarden solcher Neuronen, die in einer geeigneten Weise miteinander verbunden sind, um letztendlich das zu erreichen, was man von ihnen erwartet. Beim Deep Learning beginnt man mit einem zufällig generierten neuronalen Netz, das zunächst nicht (bzw. sehr schlecht) funktioniert. Dann werden die Trainingsdaten durch das Netz geschleust und das Ergebnis wird mit dem gewünschten Ergebnis verglichen. Anschließend werden die Verbindungen zwischen den Neuronen so angepasst, dass das Netz ein bisschen besser funktioniert. Dieser Prozess wird mit dem nächsten Trainingsbeispiel wiederholt. Am Ende funktioniert das neuronale Netz erstaunlich gut.

Vielleicht können wir zum Schluss noch mal ein bisschen auf die Geschichte des Machine Learning eingehen. Wo liegen die Anfänge des Machine Learning?

Wenn man so möchte, kann man sagen, dass Gauss mit der Methode der Linearen Regression den ersten Machine Learning Algorithmus verwendet hat. Im einfachsten Fall geht es hier darum, das Gewicht einer Person anhand ihrer Körpergröße vorherzusagen. Dies geschieht, indem eine Gerade durch die Datenpunkte gelegt wird. Die Steigung dieser Geraden gibt dann an, wie viel das Gewicht im Durchschnitt ansteigt, wenn die Körpergröße um einen Zentimeter zunimmt. Der Algorithmus versucht, die Steigung aus den Daten zu lernen, also zu bestimmen, wie die beiden Kenngrößen – Körpergröße und Gewicht – miteinander in Beziehung stehen. Bei anderen Machine-Learning-Algorithmen funktioniert dies ähnlich, nur dass man dort mit viel mehr Parametern und komplexeren Modellen arbeitet, nicht nur mit einer einfachen Geraden.

Beim Schlagwort Lernen gibt es historisch gesehen den klugen Hans. Das war ja eigentlich eine Attraktion, ein Pferd, das angeblich rechnen konnte. Was hat das mit dem Lernprozess bei Maschinen gemeinsam?

Der Fall des klugen Hans, eines Pferdes, das scheinbar Kopfrechnen konnte, ist ein interessantes Beispiel. Wenn der Mensch fragte, „Was ist acht plus vier?“, tippte das Pferd zwölfmal mit seinem Huf. Eine Untersuchung der Preußischen Akademie der Wissenschaften enthüllte jedoch, dass das Pferd nicht tatsächlich rechnete, sondern den Aufgabensteller beobachtete. Es tippte einfach mit dem Fuß, bis es die Reaktion des Menschen bemerkte, der sich entspannte, wenn das Pferd die richtige Anzahl an Schlägen getippt hatte. Das Pferd hatte also gelernt, den Menschen zu beobachten, anstatt tatsächlich zu rechnen.

Ein ähnliches Phänomen kann auch bei Machine-Learning-Modellen auftreten. Ein berühmtes Beispiel ist die Bilderkennung, bei der ein Computer lernen soll, ob ein Bild z.B. ein Hund, eine Katze oder etwas anderes zeigt. Wenn die Maschine jedoch nur zwischen Wölfen und Huskys unterscheiden soll, kann sie lernen, nach Schnee im Bild zu suchen und auf dieser

Grundlage zu entscheiden, ob es sich um einen Wolf oder einen Husky handelt. Dies kann funktionieren, ist aber nicht das, was man ursprünglich beabsichtigt hat.

Diese Algorithmen versuchen einfach, aus den Eingaben die korrekte Antwort auf den Trainingsdaten vorherzusagen, oft ohne zu verstehen, was sie tatsächlich tun. Sie können aufgrund von Zusammenhängen zwischen den Daten lernen, die nicht das sind, was man eigentlich lernen wollte, und machen dann Fehler, die nicht immer offensichtlich sind. Es ist wichtig, solche Fälle zu erkennen und zu vermeiden, um sicherzustellen, dass die Machine-Learning-Modelle das lernen, was sie sollen.

Die Fragen stellte Maximilian Weidmann.

ChatGPT, der „Tod des Autors“ und die Geltung von Texten

Lukas Ohly

1 Geltung stellt sich intersubjektiv ein

Sollen Texte die Wahrheit sagen, so müssen sie gültig sein, also Geltung haben. Dazu müssen sie selbst erst einmal *als Texte* gelten. KI-Textgeneratoren wie ChatGPT produzieren aber einfach nur sprachliche Silbenfolgen nach einem Algorithmus, der ihre wahrscheinliche Abfolge aus einem riesigen Datenmaterial errechnet. Dass der „stochastische Papagei“¹ wirklich Texte generiert, die Geltung beanspruchen können, ist daher zweifelhaft. Denn wie kann sich ein Text auf einen Gehalt gültig beziehen, wenn sich der Textgenerator auf gar keinen Gehalt bezogen hat, sondern nur eine mathematische Methode der Formel-Generierung benutzt hat, die innerhalb von Textbezügen verbleibt?

Möchte ich wissen, ob mir als Student Wohngeld zusteht oder ich vorher BAföG beantragen muss, so muss ich künftig nicht mehr kompetente Menschen befragen oder selbst im Internet nach einer passenden Antwort suchen, sondern kann auch ein sogenanntes „Large Language Model“ (LLM) befragen. Im besten Fall würde ich aus allen drei Kanälen dieselbe – nämlich richtige – Antwort dazu erhalten. Zudem können alle Quellen ihre Antwort rechtfertigen, indem sie sich auf die entsprechende Verordnung beziehen. Der Wortlaut der Verordnung wiederum kann unabhängig von den Auskünften überprüft werden, so dass ihre Geltung sichergestellt werden kann. Gültigkeit oder Geltung kann so durch eine Verweiskette von Texten bemessen werden. Da ein Sprachmodell auf ein massives Datenmaterial von Texten zurückgreift, kann es mit einer gewissen Wahrscheinlichkeit vor einem informierten Menschen bestehen.

Kann man dennoch einen Unterschied in der Geltung der Auskunft feststellen, wenn sie von einem Sprachmodell stammt? Zunächst zeigt das Beispiel, dass Geltung intersubjektiv festgestellt wird. Niemand erschafft Geltung dadurch, dass er oder sie etwas aussagt. Geltung entsteht vielmehr dadurch, dass die Aussage auf Rückfrage begründet wird und eine intersubjektiv anerkannte Verweiskette verfolgt werden kann. Wir würden dagegen einer Aussage keine Gültigkeit zumessen, wenn die Begründung dafür darin besteht, dass sie von ChatGPT stammt.

Nehmen wir an, Sie sollen aus einem Wörterpuzzle Sätze zusammenstellen, und zufällig ergibt sich bei Ihrem Puzzle der Satz: „Studentinnen und Studenten dürfen unabhängig von ihrem BAföG-Anspruch Wohngeld beantragen.“ Dann folgt nicht *daraus*, dass Sie Wohngeld beantragen dürfen, wenn Sie Studentin sind. Sie können einwenden, dass das ja auch nicht

¹ Siepmann, Dirk: Vom Akkordarbeiter zum Gutachter, 492. Bender, Emily M. u. a.: Stochastic Parrots.

der Sinn des Spiels gewesen ist. Sie haben ja nicht Sätze gepuzzelt, die schon deshalb wahr sind, weil Sie sie so puzzeln konnten, sondern die Geltung des Satzes bezieht sich in dem Spiel allein auf die Regel, grammatisch sinnvolle Sätze zu puzzeln. Ihr gepuzzelter Satz ist gültig, weil er die Regel des Spiels erfüllt. Dass er die Regel erfüllt, garantiert aber nicht der Satz selbst, sondern seine Beziehung zu den Spielregeln und zur deutschen Grammatik. Beide gehen über diesen Satz hinaus. Weder die Spielregeln noch die deutsche Grammatik legen jedoch fest, ob Studenten Anspruch auf Wohngeld haben.

Genauso unsinnig wäre es, wenn der informierte Mensch den Anspruch auf Wohngeld damit rechtfertigt, dass seine Neuronen im Gehirn den Gedanken an den Anspruch bei ihm ausgelöst haben. Zwar werden Hirnzustände zu seiner Auskunft geführt haben, aber die Geltung des Verbots liegt nicht in den Hirnzuständen dieses Menschen. Der auskunftsfähige Mensch hat zwar bestimmte Hirnzustände, wenn er auf die entsprechende staatliche Verordnung verweist. Aber ob diese Verordnung wirklich gültig ist, hängt nicht daran, dass wir beim Durchlesen dieselben Hirnzustände haben wie dieser Mensch.

Es lässt sich also die Geltung von KI-generierten Texten allein dadurch anzweifeln, dass das Rechtfertigungspotenzial von Sprachmodellen mit ihren internen Verarbeitungsmethoden nicht ausreichend ist, weil die Verarbeitungsmethoden den einzigen „Außenbezug“ der Textgenerierung bilden. Doch mit diesem Verweis ins künstliche „Gehirn“ wird noch keine äußere Verweiskette hergestellt. Die KI antwortet uns, dass Studenten Wohngeldanspruch haben, weil der „stochastische Papagei“ diese Antwort nahelegt. Der Verweis auf die entsprechende staatliche Verordnung ergibt sich wiederum aus dem „stochastischen Papagei“, aber nicht daraus, dass sie aktuell gültig ist.

Ob der KI-Text gültig ist, muss also von den Rezipienten bestätigt werden können, indem sie die behaupteten Verweise überprüfen. Zu dieser Überprüfung müssen mehr Menschen befähigt sein als nur diejenigen Instanzen, die die Auskunft erteilen. Das heißt aber dann, dass eine KI durchaus geltende Sätze erzeugen kann, weil Geltung ohnehin nicht durch die Art der Textgenerierung festgelegt wird. Geltung entsteht vielmehr intersubjektiv.

Dennoch gibt es einen Unterschied zwischen den identischen Auskünften eines informierten Menschen und einer KI: Der Mensch vertritt einen Geltungs*anspruch*, die KI nicht. Ein Geltungsanspruch besteht in der Unterstellung einer Sprecherin, dass ihre Aussage Geltung *verdient*, also zur intersubjektiven Anerkennung *verpflichtet*. Wer einen Geltungsanspruch vertritt, fordert von seinen Adressaten die Anerkennung der getroffenen Aussage. Die Sprecherin kann die Geltung nicht alleine herstellen, sondern fordert zur sozialen Kooperation auf, ihren Geltungsanspruch *gemeinsam* einzulösen. Mit dem Geltungsanspruch allein ist die Geltung der Aussage noch nicht sichergestellt. Die Sprecherin kann den Geltungsanspruch auch nicht allein einlösen, indem sie die Rechtfertigungsgründe für ihre Aussage offenlegt. Die Einlösung des Geltungsanspruchs ist erst erreicht, wenn die Aussage durch einen intersubjektiven Prozess Geltung erzielt.

Die Rechtfertigungsgründe unterstützen aber den Geltungs*anspruch*. Sie haben selbst einen Geltungsanspruch. Das würde in einen infiniten Regress führen, wenn die Sprecherin für die Einlösung des Geltungsanspruchs alleinverantwortlich wäre. Die Adressaten der Aussage

könnten dann jeden anschließenden Geltungsanspruch ohne Angabe von Gründen zurückweisen, weil die Sprecherin niemals unendlich viele Rechtfertigungsgründe abgegeben haben wird. Allerdings ist auch eine Zurückweisung von Geltungsansprüchen mit einem Geltungsanspruch verbunden. Wenn ich antworte: „Die besagte Verordnung ist nur in bestimmten Bundesländern in Kraft“, dann übernehme ich Mitverantwortung dafür, bei dem fraglichen Thema eine gültige Lösung zu erzielen. Dasselbe ist der Fall, wenn ich antworte: „Ich verstehe nicht, was du meinst.“ Wir tauschen Geltungsansprüche aus, bis sie unter uns die erforderliche Geltung erzielt haben. Natürlich könnten wir auch das Gespräch einfach abbrechen, aber dann bleiben die Geltungsansprüche solange im Raum, bis sie endgültig ausgehandelt worden sind. Jedenfalls behält die Verordnung ihren Geltungsanspruch, auch wenn wir davon nichts wissen wollen. Aber wenn sie wiederum niemand befolgt und einen Antrag auf Wohngeld stellt, hat sie auch keine Gültigkeit mehr.

Eine KI vertritt mit ihren Texten keine Geltungsansprüche. Sie fordert also von ihren Rezipienten nicht die Anerkennung ihrer Texte ein. Zwar kann sie ihre Aussagen begründen und die gleichen Verweisketten konstruieren wie ein informierter Mensch. Aber für sie selbst sind diese Verweisketten ohne Bedeutung, sondern bilden lediglich einen internen, mathematisch errechneten Zusammenhang, der nichts mit Wohngeld oder Verordnungen zu tun hat. Wenn ihre Auskunft dennoch Geltung erhält, dann durch die Rezipienten, die die Verweisketten in einem plausiblen Maß überprüft haben. Geltung entsteht aber dann, ohne dass vorher ein Sprecher einen Geltungsanspruch vertreten hätte.

Ist diese Unterscheidung valide? Oder wird wieder nur auf die unterschiedlichen internen Verarbeitungsmethoden von Aussagen bei Mensch und Maschine rekurriert, die jedoch ohnehin keine Geltungsgründe bilden? Zumindest fordern Menschen ihre Geltungsansprüche bei ihren Adressaten nicht dadurch ein, dass sie ihnen Gewalt androhen, sondern eben dadurch, dass sie Rechtfertigungsgründe anführen. Das leistet eine KI jedoch auch. Sie ähnelt der Person, die eine unendlich lange Verweiskette von Rechtfertigungsgründen bildet, wenn nicht die Nutzerin irgendwann den Prozess unterbricht und sich in den Geltungsdiskurs einmischt. Solange die Nutzerin lediglich den Diskurs stoppt, weil sie kein weiteres Interesse am Thema hat, entsteht keine Geltung. Um sich in den Geltungsdiskurs wirklich einzumischen, muss die Nutzerin irgendwann die genannten Geltungsgründe anerkennen oder Anerkennung ihrer Gegenargumente erzielen. Ein Geltungsanspruch ist so oder so mit der Forderung verbunden, dass sich die Diskurspartner *gemeinsam verantwortlich* für die Geltung der Aussage machen.

Eine KI fordert weder eine solche gemeinsame Verantwortung ein, noch kann sie sie erfüllen. Textgeneratoren demonstrieren zwar, dass Geltungsansprüche ins Unendliche iteriert werden können – einfach weil die Textgeneratoren nicht mehr aufhören. Sie haben immer das letzte Wort. Ich kann keinen „Prompt“ eingeben und dabei das letzte Wort haben, weil ich den Textgenerator nicht „überzeugen“ kann. Er kann nicht zustimmend schweigen, sondern ich kann ihn höchstens zum Schweigen bringen, indem ich aufhöre, seine Botschaft zu erwidern, oder indem ich das Programm unmittelbar nach meinem Prompt schließe. Aber das ist einem Gesprächsabbruch gleich und erzielt keine Geltung. Wäre die KI ein Mensch,

dann würde sie sich für die Geltung ihrer Aussagen alleinverantwortlich machen wollen. Aber eine KI trägt grundsätzlich keine Verantwortung. Denn sie besitzt keine persönliche Identität lässt sich daher auch nicht für ihre Texte verantwortlich machen. Sie „meint“ nicht, was sie sprachlich ausdrückt, und steht nicht hinter ihren Texten.² Daher täuscht das Wechselspiel zwischen meinen Inputs und ihren Outputs nur darüber hinweg, dass sie letztlich *allein* operiert. Sie zielt nicht auf eine gemeinsame Übereinkunft von Geltungsansprüchen, sondern erweitert ihre Verweisketten monologisch und daher prinzipiell ins Unendliche. Das Ende der Interaktion mit mir wird entweder von Programmbeschränkungen festgelegt („Du hast dein Limit für heute erreicht“) oder durch meinen Interaktionsabbruch vollzogen, aber nicht dadurch, dass dem Sprachmodell eine gemeinsam gefundene Geltung von Aussagen einleuchten würde.

2 Diskursethische Einschätzung

Mit meiner Profilierung des Ausdrucks „Geltungsanspruch“ setze ich eine andere Pointierung als die Diskursethik, wie sie etwa von Jürgen Habermas vertreten worden ist: Habermas hatte den Begriff so verstanden, dass die Sprecherin einer Aussage mit einem Geltungsanspruch selbst beansprucht ist, Gründe für sie vorzubringen. Der Geltungsanspruch muss von der Sprecherin „eingelöst“ werden.³ Auf diesem Verständnis beruht die gesamte diskursethische Pointe, dass alle, die einen Geltungsanspruch vertreten, damit eine Diskursgemeinschaft anerkennen müssen.⁴ Die Rezipienten sind allenfalls dazu beansprucht, den Geltungsanspruch als berechtigt anzuerkennen.⁵ Damit machen sie sich nicht gemeinsam mit der Sprecherin verantwortlich für die Geltung ihrer Aussage, sondern erkennen die Rolle der Sprecherin an, dass sie ihren Geltungsanspruch vertreten darf. Er ist eingelöst, sobald sie Gründe für ihn hat, nicht, wenn die Gründe allgemein geteilt werden. So kommt es strenggenommen nie zu einer gemeinsam geteilten Geltung, weil der Geltungsanspruch nie aus der Perspektive der Sprecherin zu den Rezipienten übergeht. Damit wird zugleich die Rolle der Sprecherin für den im Raum stehenden Geltungsanspruch erhöht. Ohne Sprecherin kein Geltungsanspruch!

Von der Diskursethik her müsste man nun sagen, dass Sprachmodelle keine Geltungsansprüche vertreten, weil sie keine Sprecher sind: Sie lassen sich nicht zur Einlösung von Geltungsansprüchen verpflichten. Sie sind Begründungsmaschinen, die weder ihre Adressaten als diejenigen anerkennen, die ihnen die Pflicht aufbürden, ihre Begründungen zu leisten, noch ihre Begründungen zur Einlösung von Geltungsansprüchen verstehen. Allenfalls könnte man diskursethisch etwas technikfreundlicher formulieren, dass KI zwar Geltungsansprüche

² Ohly, Lukas: Ethik für ChatGPT, 120.

³ Habermas, Jürgen: Theorie des kommunikativen Handelns Bd. 1, 424.

⁴ Habermas, Jürgen: Treffen Hegels Einwände, 11.

⁵ Habermas, Jürgen: Erkenntnis und Interesse, 389.

einlöst, aber ohne Sprecherin. Hier hängen die Geltungsansprüche in der Luft, und mit ihnen geht die diskursethische Pointe verloren, nämlich die Anerkennung der Diskursgemeinschaft.

3 Der „Tod des Autors“ und die Folgen für die Geltung

Roland Barthes hat mit seinem Ausdruck „Tod des Autors“ die Voraussetzung der diskursethischen Grundlegung von Geltung in Frage gestellt. Der Sinn einer Aussage wird nämlich bei Barthes vom Autor entkoppelt. Vielmehr entscheidet die Leserin über den Sinn.⁶ Nach gängiger Überzeugung hatte die Literaturkritik einen Text mit dem Autor erklärt:⁷ Wenn man etwa eine dunkle Stelle aus einem Paulusbrief verstehen wollte, hatte man sie mit der Biographie von Paulus erklärt. Mit Bezug auf den Autor sollten sich so möglichst alle Widersprüche aufklären lassen. Der Autor war Geltungskriterium des Textes und damit wichtiger als der Text. Diese gängige Überzeugung liegt der Diskursethik zugrunde: Ohne Autor keine Geltungsansprüche. Also können künstlich generierte Texte keine Geltungsansprüche haben, und sie können keine Geltung erzielen, es sei denn, dass sich jemand sie zu eigen macht.

Aber mit dem „Tod des Autors“ ändert sich die Situation. Nach meinem Eindruck gibt Barthes drei Hauptgründe dafür an, warum der Autor tot ist, einen historischen und zwei sachliche: Historisch hält er die Vorstellung eines Autors für eine Idee des Geniekults, wie er erst in der Moderne aufgekommen sei.⁸ Die Bibel bietet ein prominentes Gegenbeispiel für den Geniekult, weil ihre Schriften zumeist Zusammenstellungen von Redaktoren sind, also Textkollagen, deren Herkunft sich kaum bestimmen lässt. Auch die Redaktoren sind anonym.

Daraus entwickelt Barthes ein erstes Sachargument: Kein Text ist wirklich originell. Autoren mischen lediglich ihre Zitate zu einem Text, anstatt zu schreiben.⁹ Pointiert ausgedrückt, sind die Autoren selbst lediglich Sprachmodelle.

Das zweite Sachargument liegt für Barthes darin, dass Texte, die nicht in ein Geschehen eingreifen („Achtung, die Ampel da vorne ist rot!“), zum Tod des Autors führen.¹⁰ Barthes bezog sich vor allem auf Romane, die in kein reales Geschehen eingreifen, sondern ihre eigene Welt erschaffen. Nun kann man weiterfragen, ob wissenschaftliche Texte in ein Geschehen eingreifen, da sie darauf einwirken, was als wahr anzuerkennen ist, oder ob sie Romanen ähnlich sind, die zum „Tod des Autors“ führen. Für die zweite Option spricht, dass bei der aktuellen unermesslichen Menge wissenschaftlicher Texte nur wenige Publikationen wirklich Einfluss auf eine reale Veränderung haben.

⁶ Barthes, Roland: Der Tod des Autors, 63.

⁷ Ebd., 58.

⁸ Ebd., 57.

⁹ Ebd., 63.

¹⁰ Ebd., 57.

Mit dem Tod des Autors kommt es zur „Geburt des Lesers“. Der Leser ist von seiner Geschichtslosigkeit zu befreien. Ein „Mensch ohne Geschichte“¹¹ ist er, solange es gleichgültig ist, ob ich den Text lese oder Sie. Es müsste dann nämlich immer dasselbe Interpretationsergebnis herauskommen, weil jeder Leser stets der Autorenabsicht folgt, die sich im Text eindeutig niedergeschlagen hätte. Und da nicht der Text die Richtgröße wäre, sondern die Autorenabsicht, würde der Text sie nur repräsentieren.

Wenn dagegen der Leser ein Mensch *mit* Geschichte ist, verändert sich mit ihm der Text. Dadurch vervielfacht sich der ursprüngliche Text. Im ersten Fall liegt die Geltung hinter dem Text, im zweiten Fall in der Vervielfältigung des Textes durch die Leser. In beiden Fällen liegt der Sinn nie im Text selbst.

Michel Foucault folgt Barthes' Einschätzung, dass Texte in der Moderne mit dem Autor erklärt wurden.¹² Für Foucault wird der „Tod des Autors“ zum einen ein politisches Befreiungsmittel, das sich gegen eine Kultur stellt, die Gedanken zum Privateigentum erklärt.¹³ Der Autor wird in dieser Kultur als eine gesellschaftliche Funktion benutzt.¹⁴ Diese Funktion besteht darin, Diskurse zu reglementieren.¹⁵ Durch den Autor werden bestimmte Textbeiträge ausgeschlossen. Er fungiert als Verknappung von Diskursen.¹⁶ Wer sich etwa kritisch mit der Psychoanalyse Freuds auseinandersetze, müsse bestimmte Leitsätze Freuds als unhintergebar akzeptieren. Jegliche Interpretation stehe im Horizont dieser Leittexte. Wer dagegen diese Leittexte kritisch zurückweisen wolle, könne nicht mehr über Freud schreiben. Also bleibt sogar Kritik an Freud eingefangen im System dieses Autors: „Der Begründungsakt eines wissenschaftlichen Fachs kann im Zuge der Weiterentwicklung dieser Wissenschaft nur wie ein Sonderfall in einem viel allgemeineren Ganzen erscheinen.“¹⁷ Diesem Verknappungssystem von Diskursen kann man nur durch den „Tod des Autors“ entkommen.

In diesem Verknappungssystem ist der Autor kein Individuum, sondern ein soziales Herrschaftssystem. Er wird als eine „Doktrin“ benutzt: „Die Doktrin führt eine zweifache Unterwerfung herbei: die Unterwerfung der sprechenden Subjekte unter die Diskurse und die Unterwerfung der Diskurse unter die Gruppe der sprechenden Individuen.“¹⁸ Das klingt paradox, weil sich hier beide, sprechende Individuen und Diskurse, gegenseitig unterwerfen. Aber die „sprechenden Individuen“ sind eben in der Doktrin keine Subjekte, sondern eine „Gruppe“. Der Autor wird durch mächtige Gruppen – nämlich Verlage,¹⁹ privatrechtliche

¹¹ Ebd., 63.

¹² Foucault, Michel: Was ist ein Autor?, 213.

¹³ Ebd., 205.

¹⁴ Foucault, Michel: Die Ordnung des Diskurses, 20.

¹⁵ Foucault, Michel: Was ist ein Autor?, 211, 228.

¹⁶ Foucault, Michel: Die Ordnung des Diskurses, 36.

¹⁷ Foucault, Michel: Was ist ein Autor?, 221.

¹⁸ Foucault, Michel: Die Ordnung des Diskurses, 30.

¹⁹ Ebd., 28.

Institutionen (Autorenrechte²⁰) zur Sicherung des Privateigentums²¹ – konstruiert mit der Funktion der Diskursverknappung.

Nun könnte man meinen, dass Foucault die sprechenden Individuen vor dieser gesellschaftlichen Bemächtigung retten will. Dann müsste er eigentlich die Autoren vom Zugriff eines herrschenden Systems retten wollen. Das ist aber nicht der Fall, denn jede Rettung des Autors führt ja zu eben der Verknappung der Diskurse, die ihm seine Individualität nimmt. Vielmehr liegt die Lösung darin, dass man andere Fragen an Texte legt: „Folgende so lange wiedergekäute Fragen würde man nicht mehr hören: ‚Wer hat eigentlich gesprochen? Ist das auch er und kein anderer? ...‘ Dafür wird man andere hören: ‚Welche Existenzbedingungen hat dieser Diskurs? Und woher kommt er? Wie kann er sich verbreiten, wer kann ihn sich aneignen? Wie sind die Stellen für mögliche Stoffe verteilt?‘“²²

Foucault beschreibt also den „Tod des Autors“ als ein Dilemma, weil selbst die Befreiung aus dem Herrschaftssystem der Interpretation nicht zur Rückgewinnung des souveränen Subjekts führt. Befreien lässt sich allenfalls das literarische Werk vom „Zwangssystem“²³ der Literatur, aber nicht indem damit der Autor wieder zur Sprache kommt, sondern indem sich die Interpretation von der Deutungsmacht der herrschenden Meinung emanzipiert und kreative Wege geht: „[Die Literaturkritik] löst sich von den alten Themen der Einschachtelung, der Schatztruhe, die man im Werk suchen und öffnen muss. Sie stellt sich außerhalb des Werkes, schafft so ein neues Außen und schreibt Texte aus Texten.“²⁴

Ich lese Foucaults Diktum vom Tod des Autors im breiten Kontext seiner Arbeiten, die vor allem die gesellschaftliche Aus- und Einschließung der „Wahnsinnigen“ in den Blick nahm, der sozialen Macht das Konstrukt der „Wahrheit“ zurechnete und daraus eine Ablösung der Subjektivität folgte. Der Autor ist dann vor allem deshalb tot, weil er kein Subjekt ist: Er ist kein Subjekt, weil ein solches als „Scheinsubjekt“²⁵ sozialer Macht konstruiert wird. Seine Befreiung muss dann auch den Subjektstatus bekämpfen.²⁶ Foucault hat seine eigenen Texte so verstanden, dass er sich mit ihnen nicht auf sich festlegen wollte: „Ich denke niemals völlig das Gleiche, weil meine Bücher für mich Erfahrungen sind.“²⁷ Ebenso fordert er, dass wir aus uns „Kunstwerke“ machen, die mit „der schöpferischen Praxis und nicht mit der Idee der Authentizität“²⁸ zu tun haben. Wenn die Wahrheit eigener Texte in der Zukunft liegt,²⁹ dann

²⁰ Foucault, Michel: Was ist ein Autor?, 212.

²¹ Ebd., 211, 229.

²² Ebd., 227.

²³ Foucault, Michel: Schriften 3, 331.

²⁴ Foucault, Michel: Schriften 1, 759.

²⁵ Foucault, Michel: Schriften 2, 277.

²⁶ Ebd.

²⁷ Foucault, Michel: Schriften 4, 52.

²⁸ Ebd., 474.

²⁹ Ebd., 51.

ist sie dem Subjekt und der es konstituierenden sozialen Macht entzogen und wird zur Sache der „Wahrnehmung ... bei den Menschen“³⁰.

Der Ansatzpunkt, dass die Emanzipation von der Deutungsmacht eines Zwangssystems nur durch die Dekonstruktion des Subjekts gelingen kann, liegt in Foucaults Analysen der Geschichte des Wahnsinns. Eine zentrale Zäsur liegt dabei in Sigmund Freuds Psychoanalyse, der es nämlich darauf ankommt, im psychisch Kranken gerade durch das Aussprechen des Wahnsinns an die Wahrheit seines psychischen Zustandes heranzukommen. Gerade also weil der Wahnsinn ausgesprochen wird, enthüllt sich darin die Wahrheit. Das bedeutet, dass die Wahrheit nicht darin liegt, *was* jemand sagt.

Seit Freud ist der abendländische Wahnsinn zu einer Nicht-Sprache, weil zu einer doppelten Sprache geworden (eine Sprache, die allein in diesem Sprechen existiert; ein Sprechen, das nichts als seine Sprache sagt) –, das heißt eine Matrix der Sprache, die, im strengen Sinne, nichts sagt. Eine Falte des Gesprochenen, die eine Abwesenheit des Werkes ist.³¹

Wichtig ist dabei, dass die „Wahrheit“ unabhängig vom Willen des Sprechenden ausgesprochen wird.³² Der Sprecher ist also kein Souverän. Indem vielmehr die Wahrheit des Wahnsinns durch eine therapeutische Technik geborgen wird, steht sie unter dem Vorzeichen sozialer Macht. Diesen Mechanismus verallgemeinert Foucault, indem er Diskurstechniken aufspürt, denen Menschen überhaupt unterworfen sind.

Welch bedeutender Moment, bei dem eine Gesellschaft der anonymen Masse der Leute Worte, Wendungen und Sätze, Sprachrituale lieh, damit sie über sich selbst sprechen können – öffentlich und unter der dreifachen Bedingung, dass dieser Diskurs in einem klar definierten Machtdispositiv adressiert und in Umlauf gebracht wird, dass er den bis dahin kaum wahrnehmbaren Grund der Existenzen erscheinen lässt und dass er ausgehend von diesem winzigen Krieg der Leidenschaften und der Interessen der Macht die Möglichkeit zu einer souveränen Intervention gibt.³³

Von nun her wird verständlich, dass zwischen „ich lüge“ und „ich schreibe“ kein prinzipieller Unterschied besteht,³⁴ weil es sich in beiden Fällen um einen performativ widersprüchlichen Sprechakt handelt.³⁵ Er bezeichnet einen „leeren Selbstbezug“³⁶ des Subjekts und belegt damit, dass es nicht existiert, sondern nur in der Klammer sozialer Macht auftritt.

³⁰ Ebd.

³¹ Foucault, Michel: Schriften 1, 547.

³² Foucault, Michel: Schriften 2, 137.

³³ Foucault, Michel: Schriften 3, 329.

³⁴ Foucault, Michel: Schriften 1, 549.

³⁵ Foucault, Michel: Schriften 4, 207f.

³⁶ Foucault, Michel: Schriften 1, 549.

Foucaults Gedanken heben damit die Geltungsansprüche des Autors auf. Zwar mögen sie vom „Zwangssystem“ stilisiert worden sein, werden aber gerade deshalb auch schon wieder zurückgenommen, eben weil sie aus dem Zwangssystem heraus konzipiert sind. Die Geltung eines Textes liegt nicht in der Autorenintention, sondern im Machtapparat von Verlagen³⁷ und Institutionen.³⁸ Man kann sich von dieser Macht nur emanzipieren, indem man Geltung *aufföst*, indem man sie kreativ-kombinatorisch unterläuft oder der Zukunft überlässt. An die Stelle von Geltungsansprüchen tritt das In-Erscheinung-Bringen, das aber damit nur durch die Wahrnehmenden wahr wird.

4 Sind Sprachmodelle Module zur Befreiung?

Nun bemächtigen sich auch KI-Sprachmodelle der oder vielleicht sogar aller öffentlichen Diskurse. Zwar steht hinter einem GPT kein Autor mehr, aber die „Existenzbedingungen“ des Diskurses sind auch nicht frei, sondern durch die jeweiligen Algorithmen, ihrem Datenmaterial und ihrer Datenverarbeitung beschränkt – und damit letztlich durch die Privatinteressen seiner Anbieter. Könnte also nicht gerade durch den Tod des Autors die Macht des „Herrschaftswissens“ nochmals gestärkt werden, weil weder die Algorithmen den Nutzern bekannt sind noch ihre Datengrundlagen, mit denen sie prozessieren?

Interessanterweise scheint es aber so zu sein, dass Foucault der Entwicklung von KI-Textgeneratoren aufgeschlossen gewesen wäre. Diskurse mit den Mitteln der Datenverarbeitung zu analysieren, schien ihm ein verheißungsvoller Weg zu sein, um die Deutungsmacht positivistischer Wissenschaften zu brechen.³⁹ Die neue Kombinatorik der Literaturkritik, die Foucault anpreist, könnte auch Algorithmen umfassen, die damit dem Ziel dienen, „ein neues Außen“ zu schaffen und „Texte aus Texten“ zu schreiben.⁴⁰ Schließlich sympathisiert Foucault mit der These von Lévy-Strauss, die strukturelle Linguistik könne ihre Wissenschaftlichkeit durch „Techniken der Formalisierung“ stärken, die auch „Übersetzungsmaschinen“ einschließe.⁴¹ „Die strukturelle Linguistik zielt nicht auf empirische Ansammlungen individualisierbarer Atome ..., sondern auf systematische Gesamtheiten von Beziehungen zwischen Elementen.“⁴² Eine solche „logische Analyse des Realen“⁴³ erzeugt eben eine neue

³⁷ Foucault, Michel: Schriften 2, 146.

³⁸ Foucault, Michel: Schriften 4, 274f, 897.

³⁹ Foucault, Michel: Schriften 1, 722.

⁴⁰ Ebd., 759.

⁴¹ Ebd., 1042.

⁴² Ebd., 1045.

⁴³ Ebd., 1047.

Kombinatorik von Beziehungen „unabhängig von den Elementen, auf die sie sich beziehen“⁴⁴, und löst sich von einer Theorie des Subjekts zugunsten „sämtlicher Phänomene der Information“.⁴⁵

In diesen Erwartungen geht Foucault jedoch anscheinend davon aus, dass die „logische Analyse des Realen“ transparent ist – sei es, dass die Realität transparent und überprüfbar ist, sei es, dass die Verfahren methodisch nachvollzogen werden können. Er unterstellt also einen Wissenschaftsbetrieb, der sich in der Macht eines „Zwangssystems“ relativ frei bewegen kann. Dazu müssten die Verfahren also nicht nur den Wissenschaften zur Verfügung stehen, sondern auch die Ergebnisse einer *freien* Kombinatorik dürften nicht durch diese Verfahren beschränkt werden. Solche Beschränkungen liegen aber vor. Anstatt die Programmcodes der Öffentlichkeit zur Verfügung zu stellen, müssen gegenwärtig die Muster der Textgenerierung von Sprachmodellen mühsam erforscht werden.⁴⁶ Eine „logische Analyse des Realen“ müsste sich daher auch auf die Algorithmen der Sprachmodelle erstrecken.

Dieser Schluss kann nach meinem Eindruck nur unter zwei Bedingungen umgangen werden:

1. Entweder wird die logische Analyse von Algorithmen ihrerseits von Algorithmen durchgeführt, die aber für die Wissenschaftsgemeinde transparent sind.
2. Oder es ist gleichgültig, wer die Kombinatorik von Textbeziehungen vornimmt, weil die Kombinatorik nicht selbst die Wahrheit des „Realen“ aufspürt, sondern lediglich den Raum eröffnet, in dem frei weiterinterpretiert und -kombiniert werden kann.

In beiden Fällen liegt das Befreiungsmoment der Literaturkritik darin, dass *Menschen* die neuen Beziehungen herstellen oder nachvollziehen können. Ansonsten bleiben Geltungsansprüche an das System „Autor“ gebunden – und sei dieser Autor auch ein Digitalkonzern und sein Text ein Algorithmus, ein „Text aus Texten“. Nach meinem Eindruck also kann Foucaults emanzipatorisches Interesse an der Dekonstruktion von Geltungsansprüchen die neuen digitalen Machttechniken nicht unterlaufen, wenn Geltung lediglich in die Zukunft verschoben werden soll, über die ein „Zwangssystem“ nicht verfügen kann. Vielmehr muss ein präsentischer Geltungsbereich bestehen bleiben, der sich darauf bezieht, dass *Menschen* mit Texten operieren.

Ich nehme also Foucaults Analyse von Deutungshoheiten im Hinblick auf KI-Textgeneratoren kritischer als Foucault in seinen Andeutungen selbst. Versteht man nämlich Algorithmen selbst als Texte, die aufgrund ihrer Intransparenz – auch indem sie sich durch selbstlernende KI-Systeme selbst fortschreiben – nicht ohne zirkulären Selbstverweis kritisierbar sind, so sind sie selbst „Autoren“, die durch ein System der Autorenschaft in Geltung gesetzt

⁴⁴ Ebd., 1045.

⁴⁵ Ebd., 1047.

⁴⁶ Z. B. durch Experimente mit Generative AI (Meier-Vieracker, Simon: Uncreative Writing, 141. Hiltmann, Torsten: Hermeneutik in Zeiten der KI, 212ff. Stöcker: Lernende Maschinen und die Zukunft der Öffentlichkeit, 412. Stiemer, Haimo / Gius, Evelyn / Gerstorfer, Dominik: Künstliche Intelligenz und literaturwissenschaftliche Expertise, 458f.).

werden. Was Foucault im Hinblick auf die zirkuläre „Diskursivitätsbegründung“⁴⁷ von Autor und Werk aussagt, lässt sich dann auch auf KI-Textgeneratoren übertragen: „Folge davon ist, daß man die theoretische Gültigkeit in bezug auf das Werk dieser Begründer selbst definiert.“⁴⁸ Die Geltung eines KI-Textgenerators also dadurch erzeugt, dass sie sich diesem Textgenerator als *Autor* verdankt, aber andererseits zu seinem *Werk* gehört. Als Diskursivitätsbegründung wird das Verfahren des KI-Textgenerators nicht etwa deshalb gültig, weil es wahr oder hinreichend verifiziert worden wäre. Vielmehr wird es für gültig erklärt, weil es das *Werk* des KI-Textgenerators ist. Autor und Werk setzen sich in einem Zirkelschluss wechselseitig in Geltung. Was Foucault hier „selbst definiert“ nennt, meint diesen Zirkelschluss. Ebenso wie die Gültigkeit des Werks eines „Diskursbegründers“ (zum Beispiel die Namen Freud oder Marx⁴⁹) keine „Außenfenster“ zur nicht-literarischen Wirklichkeit hat, beziehen sich zum Ersten auch die Texte von GPT nur auf sich selbst und auf ihren internen Wahrscheinlichkeitsabgleich. Zum Zweiten aber sind ihre auch Algorithmen unhintergehbare „Diskursbegründer“. Mit dem künstlich intelligenten Tod des Autors entsteht die Quasi-Neugeburt eines Ersatzsystems, das das System „Autor“ nach dessen Tod ablöst.

Foucaults Kritik am System des Autors muss sich daher nach meinem Eindruck auch auf Texte beziehen, die von einer „halluzinierenden KI“⁵⁰ geschrieben sind. Zwar kann uns kein Autor mehr hinters Licht führen, weil ohnehin nicht er über seine Geltung entscheidet. Roland Barthes gibt als Beispiel dafür das surrealistische Schreiben, das dem Textgenerieren einer KI nahe kommt: Hier schreibt jemand schneller, als der Kopf denkt.⁵¹ Deshalb kommt nichts mehr zustande, was die schreibende Person „meinen“ kann. Das heißt aber nicht, dass die Macht verschwindet, die Geltung etabliert. Es spielt dafür keine Rolle, dass dafür der Autor nicht einmal mehr sterben muss, sondern der Text keinen Autor hatte.⁵² Foucault kann also Sprachmodelle mit seinem Emanzipationsinteresse gegen die machtförmige Entstehung von Geltung nicht zu Hilfe nehmen. Er muss voraussetzen, dass die Emanzipationsbewegung letztlich von Menschen ausgeht.

5 Leblose KI und lebendige Leser

An die Stelle der Geltung von Texten tritt damit die Geltung freier Übereinkunft von Diskursmitgliedern. Auch wenn sich Foucault kritisch gegen Habermas' „utopische“⁵³ Vorstellung

⁴⁷ Foucault, Michel: Was ist ein Autor?, 223f.

⁴⁸ Foucault, Michel: Schriften 1, 722.

⁴⁹ Foucault, Michel: Was ist ein Autor?, 219.

⁵⁰ Bahr, Amrei: Same same but different, 180f. Der Begriff des halluzinierenden Papageis stammt von Bender, Emily M. u. a.: Stochastic Parrots, 617.

⁵¹ Barthes, Roland: Der Tod des Autors, 59.

⁵² Winter-Tietel, Roman: Wenn Niemand Texte schreibt, 268f.

⁵³ Foucault, Michel: Schriften 4, 898.

von Kommunikationsbeziehungen äußert, bei denen der Konsens einen herrschaftsfreien Bereich etabliert,⁵⁴ scheint der Hebel in Foucaults Befreiungsemphase darin zu liegen, dass eine Diskursgemeinschaft Geltungsansprüche in Texte *hineinliest*. Zumindest wenn sie sie als Texte auffassen, können die Leser nicht anders, als sie so zu lesen, dass sie etwas meinen.

Das nun trifft aus meiner Sicht auch auf Texte zu, die von sich aus keine Geltung haben können, weil sie von einer „halluzinierenden KI“ geschrieben sind. Denn wenn man ihnen unmittelbar keine Geltung und keinen Gehalt unterstellt, wären sie schlicht funktionslos: Man könnte nicht entdecken, dass in ihnen etwas steht. Das heißt gerade nicht, dass Menschen KI-generierten Texten absichtlich einen Geltungsanspruch unterstellen, wissen sie doch eigentlich, dass sie aus einem rein formalen Verfahren zustande kommen. Und es heißt auch nicht, dass Menschen solchen Texten nur diejenigen Geltungsansprüche abgewinnen, die sie selbst vertreten. Denn ansonsten wüssten die Leser vorher bereits, was in den Texten steht, und müssten sie nicht erst noch lesen. Auch in diesem Fall hätten KI-Textgeneratoren keine Funktion.

Vielmehr *widerfährt* den lesenden Menschen die Geltungsebene unwillkürlich, es sei denn, sie fassen Texte nicht einmal als Texte auf, sondern etwa als ästhetische Zeichenfolgen mit bloß dekorativer Funktion. Aber schon wenn sie als verschlüsselte Codes verstanden werden, den die Leser selbst nicht knacken können, ahnen Menschen darin einen Geltungssinn. Das Widerfahrnis der Geltungsebene ereignet sich also unabhängig von einem Autor. Ich erweitere also mit meinem Vorschlag Foucaults These vom Tod des Autors um den Widerfahrnischarakter von Geltung. Denn mit Foucaults Dekonstruktion des Autors verschwindet eben nicht das System der Geltung. Und es verschwindet vor allem deshalb nicht, weil sich jegliche Systemmacht das Widerfahrnis der Geltung zunutze machen kann.

Welchen Geltungsanspruch der Text setzt, ist aber nicht ebenso durch ein Widerfahrnis gegeben, sondern muss von den Lesern ermittelt werden. Dasselbe trifft auf die Überprüfung zu, ob der Text den Geltungsanspruch zu Recht enthält, also ob er wahr ist. Darüber entscheidet nun nicht die Willkür des Lesers, sondern die sozial vermittelte Geltungsfindung. Das ist einerseits der Ort der Deutungsmacht eines Zwangssystems, aber andererseits auch der Ort, durch freie Kombinatorik der sprachlichen Beziehungselemente dieses System zu unterlaufen. Allerdings lässt es sich aus meiner Sicht nicht unterlaufen, ohne dabei unmittelbar auf Geltung aus zu sein.

Die Geltungssuche erfordert aber ein gemeinsames Vertrauen der Leser ineinander. Fehlt sie, könnten sich die Leser nicht einmal darauf einigen, es mit demselben Text zu tun zu haben. Die Geltungssuche widerfährt aber nicht ebenso unvermittelt wie die Entdeckung der Geltungsebene eines KI-generierten Textes. Die Paradoxie besteht ja nach wie vor, dass die Leser wissen (können), dass Sprachmodelle keinen Geltungsanspruch setzen, dass man ihnen aber nicht anders begegnen kann als mit der unvermittelten Unterstellung, dass diese Texte eine Geltungsebene besitzen. Diese Paradoxie macht es denkbar, dass andere lesende Menschen die Geltungsansprüche desselben Textkorpus willkürlich setzen. Das vertieft die

⁵⁴ Ebd., 721.

Paradoxie: Obwohl ich den fordernden Charakter des Geltungsanspruchs akzeptiere, den der Text an sich nicht hat, der mir aber widerfährt, unterstelle ich den mitlesenden Menschen die Möglichkeit, dass sie diesen fordernden Charakter nicht akzeptieren, weil sie – zu Recht – durchschauen, dass der Text keinen Geltungsanspruch hat.

Die Lösung aus dieser Paradoxie besteht nach meinem Eindruck darin, dass das Widerfahrnis der Geltungsebene mit seinem fordernden Charakter *allgemein* ist und eine intersubjektive Verbindlichkeit bei der Geltungssuche auslöst. Dafür gibt es zunächst keinen anderen Grund als das Widerfahrnis der Geltungsebene selbst. Alle beteiligten Diskurspartnerinnen und -partner müssen dazu einander *vertrauen*, dass sie ebenso ernsthaft an der Geltungssuche interessiert sind, und zwar obwohl sie gute Gründe hätten, Geltung von Texten zurückzuweisen, die eine „halluzinierende KI“ geschrieben hat. Das wechselseitige Vertrauen der Leser ergibt sich also nicht ebenso unmittelbar wie das Widerfahrnis der Geltungsebene. Und doch hat die unmittelbare Entstehung der Geltungsebene ein stärkeres Gewicht als die Zweifel an der Ernsthaftigkeit der übrigen Leser, weil die Geltungsebene unwillkürlich widerfährt und es uns andernfalls nicht einmal so erscheinen könnte, dass wir uns gemeinsam mit dem Textsinn auseinandersetzen.

In diesem günstigen Fall also vertrauen die beteiligten Leser einander, dass der Text sie alle verpflichtet, seinen Geltungsanspruch zu überprüfen. Das heißt nicht, dass sie sich darauf einigen können, welche Geltungsansprüche der Text setzt. Aber bereits das Vertrauen, auf Geltung aus zu sein, hat Geltung und fordert seine Wechselseitigkeit. Dieses Vertrauen bezieht sich auf Texte, und da die Leser dem fordernden Charakter des wechselseitigen Vertrauens nicht entkommen, schafft es eine gültige Basis für das Textverstehen. Zwar mag unter den Lesern Konsens sein, dass der Autor tot ist, zumal sich die Paradoxie KI-generierter Texte durchschauen lässt, dass hinter ihnen niemand steht, der einen Geltungsanspruch mit diesem Text vertritt. Dennoch ringen sie gemeinsam um seine Gehalte, weil seine Geltung widerfährt und einen unmittelbaren fordernden Charakter hat, seine Geltung zu überprüfen.

Allerdings ist dieses Vertrauen ineinander auch nicht selbstverständlich, weil es sich nicht unmittelbar ergibt, sondern mit einer paradoxen Situation einhergeht. Deshalb ergeben sich unterschiedliche Hürden, die Geltung eines Textes gemeinsam zu bestätigen. Vergleichen wir dazu verschiedene Fälle.

1. Im ersten Fall lesen wir gemeinsam einen von einem Menschen geschriebenen Text. Ich vertraue darauf, dass Sie denselben Text wie ich mit derselben Ernsthaftigkeit verstehen wollen, um herauszufinden, ob seine Gehalte gültig sind. Aber das Umgekehrte ist nicht der Fall: Sie unterstellen mir, dass ich den Text gar nicht verstehen will, sondern ihn nur instrumentalisiere, um bestimmte Machtvorteile zu erreichen. Sie vermuten, dass ich den Text für meine Zwecke absichtlich missverstehe. Ich lese nach Ihrem Verdacht also nicht wirklich den Text, sondern kaschiere meine Interessen hinter ihm. In dieser Situation wird sich keine Geltung des Textes ermitteln lassen, die wir gemeinsam teilen. Weil ich Ihnen vertraue, werde ich zwar voreilig diejenigen Textinterpretationen für gültig erklären, die wir miteinander teilen. Aber da Sie mir umgekehrt misstrauen, können Sie mir schon hierin nicht mehr beipflichten: Für Sie besteht nämlich nur der Schein derselben Textinterpretation, weil der Text ja etwas anderes bedeuten muss, wenn die Interpretation von mir stammt.
2. Textgeltung kann also nach dem Tod des Autors nur entstehen, wenn wir uns wechselseitig vertrauen, dass wir denselben Text mit derselben Achtsamkeit interpretieren. In diesem zweiten Fall besprechen wir zum Beispiel einen Bibeltext und vertrauen darauf, dass es uns ernst ist, ihn verstehen zu wollen. Vielleicht einigen wir uns sogar auf gemeinsame Interpretationsmethoden. Obwohl wir verschiedenen religiösen Milieus zugehören oder teilweise sogar aus unreligiösen Menschen bestehen, erkennen wir einander an, kompetente und befugte Interpreten zu sein. Hier ist es unerheblich, wenn Sie keine Christin sind oder die historisch-kritische Methode nicht akzeptieren. Denn der Text kann mit dem Tod des Autors nicht beanspruchen, dass er nur von Christen oder nur mit einer einzigen Methode gelesen werden darf – das wäre eine nicht-textliche Voraussetzung, die etwas über uns aussagt, aber nichts über den Text. Bei diesem zweiten Fall aber verhält sich niemand so überheblich, die anderen Leser für inkompetent zu halten.

Ich behaupte nicht, dass dieser zweite Fall häufig vorkommt, aber dass er immer wichtiger wird in Zeiten, in denen sich die Textwissenschaften dem Tod des Autors angeschlossen haben⁵⁵ und KI Texte generiert. Das trifft auf die nun folgenden Fälle zu.

3. Im dritten Fall lesen Menschen einen KI-generierten Text, beispielsweise Studierende und ihre Dozentin eine KI-Abhandlung über ein seminarrelevantes Thema. Wenn nun eine KI keinen Geltungsanspruch vertritt und mit dem Tod des Autors ohnehin keine Geltungsansprüche mehr *im Text* liegen, dann kann Geltung nur in der Seminarsitzung „geboren“ werden. Einzelne Studierende oder die Dozentin nehmen dabei diesen Text für sich in Anspruch, der selbst nichts behauptet, da hinter ihm kein Autor steht. Das ist eine günstige Situation, die in den zweiten Fall führt: Weil sich die Leser vertrauen, kann Geltung entstehen. Sie entsteht aber nicht, weil im KI-Text Wahrheit beansprucht ist, sondern weil eine Interpretationsgemeinschaft die Geltung erst

⁵⁵ Besenbäck, Raphael / Prager, Lorenz: 502. Peiter, Anne D.: 273.

erschafft, die sie dann rückwirkend im Text lokalisiert.⁵⁶ Dieses Verfahren ähnelt der Besprechung von Naturgegenständen, denen auch keine Geltung innewohnt, sondern die wir in sie hineinragen (zum Beispiel durch Lehrsätze oder Naturgesetze).

4. Im letzten Fall lesen Menschen einen Text und verwenden dabei Sprachmodelle als „Gesprächspartner“. Bei diesem Fall spielt es keine Rolle, wer den Quelltext geschrieben hat. Hier entsteht keine Geltung, weil kein wechselseitiges Vertrauen in die ernsthafte Lesehaltung aller Beteiligten entstehen kann. Der Mensch kann zwar in die Technik eines KI-Textgenerators vertrauen, so wie er der Funktionsfähigkeit einer Spülmaschine vertraut. Ein unbedarfter Mensch wird sogar in die Lesekompetenz der KI vertrauen, obwohl sie lediglich Texte nach einem Algorithmus „durchsucht“, aber nicht in einer Sinnrichtung erfasst.⁵⁷ Aber die KI vertraut nicht in den Menschen; sie monologisiert so lange, bis der Mensch den Kontakt abbricht. Sie kann zwar den vorgegebenen Text zusammenfassen, gliedern, paraphrasieren und gewichten, aber sie kann dabei nie Geltung erzielen. Denn Geltung entsteht nur gemeinsam. Dieser Fall mündet damit in den ersten: Ohne wechselseitiges Vertrauen kann keine Geltung erzielt werden. So „zeichnet sich eine Krise der Quellenkritik ab.“⁵⁸

Während nun der zweite Fall in Zeiten KI-generierter Texte immer wichtiger wird, belegen die anderen Fälle, dass dieses Vertrauen nicht selbstverständlich ist und auch nicht einfach unbegründet vorliegt: Wir müssen uns vielmehr dieses Vertrauen verdienen. Auch dahinter steht wieder ein Geltungsanspruch, der nur gemeinsam eingelöst werden kann, indem wir uns wechselseitig für die Geltung dieses Vertrauens verantwortlich machen. Das setzt *zwischenmenschliche* Kommunikation voraus – und damit die Erwartung, dass mit uns wirklich ein Mensch kommuniziert und keine KI. Je stärker KI-Textgeneratoren die zwischenmenschliche Kommunikation allein durch eine Textüberflutung in die Defensive drängen, desto weniger Raum besteht für dieses wechselseitige Vertrauen. Wenn eine Studentin an die Dozentin eine E-Mail schreibt, in der sie über einen Gedanken aus einer Vorlesung diskutieren will, aber eine merkwürdige Antwortmail erhält, in der weitschweifig, nicht festgelegt und wenig spezifisch auf ihr Anliegen reagiert wird, wird sie sich fragen, ob beide überhaupt dasselbe Anliegen haben – ob so also noch Geltung entstehen kann. Ebenfalls ärgern wir uns über Beschwerdeportale, bei der wir uns mit einem Sprachcomputer austauschen müssen, bis wir vielleicht irgendwann an eine Person am Telefon weitergeleitet werden, die unser Anliegen versteht. Hier scheint die Metapher vom Tod des Autors an eine Grenze zu stoßen, und zwar auch dann, wenn die Geltung nicht im Text liegt. Wir brauchen Menschen, denen wir vertrauen, dass sie mit uns gemeinsam Geltungsbereiche ausloten.

⁵⁶ Peiter, Anne D.: KI-Texte = Intertexte?, 273f. Besenbäck, Raphael / Prager, Lorenz: Künstliche-Intelligenz-Quellen, 502.

⁵⁷ Ohly, Lukas: Ethik für ChatGPT, 26.

⁵⁸ Besenbäck, Raphael / Prager, Lorenz: Künstliche-Intelligenz-Quellen, 519.

6 Ethische Folgerungen

„Wird ein LLM zum Ko-Autor, verschwindet der Mensch insofern ein Stück weiter, da nicht mehr sicher ist, ob bestimmte diskursive Aussagen der Person oder dem Modell zuzuordnen sind.“⁵⁹ Je stärker die zwischenmenschliche Kommunikation durch die Flut geltungsloser Texte und vertrauensloser Instanzen verstopft wird, desto mehr beschränkt sich das Textverstehen auf subjektive Eindrücke ohne soziale Geltung. Umso wichtiger wird es, dass wir die Bereiche kennen, in denen eine KI mit uns Texte austauscht und in denen wir mit Menschen kommunizieren. Wo Menschen jedoch nur noch ausnahmsweise miteinander über Texte kommunizieren, werden Geltungsfragen aus der sprachlichen Interaktion weitgehend herausgehalten. Das ist ein ethischer Verlust: Nicht nur die Prinzipien der Diskursethik, die aus Geltungsansprüchen die Achtung der Menschenrechte rekonstruieren kann,⁶⁰ würden so umgangen werden. Es würde darüber hinaus auch unsere Empfindsamkeit für Geltungsfragen gestört. Wir würden Texte hinnehmen als Träger von Daten, die weiter zirkulieren müssen, damit sie etwas bewirken. „Bewirken“ wäre dabei eine physische Eigenschaft – unabhängig davon, ob die Informationen „wahr“ sind oder Geltung verdienen. Das wäre auch für Foucault keine wünschenswerte Situation, weil er die logische Formalisierung von Texten gerade vom „Positivismus“ der Kausalität befreien wollte.⁶¹

Wenn wir mit KI-Textgeneratoren den Sinn eines Textes ermitteln wollen, so wird dabei keine Geltung der menschlich-künstlichen Interpretation entstehen. Möglich ist es aber, dass durch den zwischenmenschlichen Diskurs Geltung in KI-Texte hineingelegt wird. Das setzt eine Kultivierung zwischenmenschlicher Kommunikation voraus, bei der die Kommunikationspartner das wechselseitige Vertrauen bestätigen können, ernsthaft zu interpretieren. Diese Kultur muss sogar noch intensiviert werden, je mehr KI-Texte im Umlauf sind.

Bei dieser Kultivierung können die Kirchen eine wesentliche ethische Aufgabe übernehmen, weil sie in ihrem Wesen Interpretationsgemeinschaften sind und weil sie den Widerfahrnscharakter theologischer Geltung eigens in den Blick nehmen. Bei allen konfessionellen Unterschieden stellen sich nämlich aus theologischer Sicht Geltungen im synodalen Prozess ein. Umso wichtiger ist es auch in den Kirchen, der Versuchung zu widerstehen, angesichts von Kostendruck und zurückgehenden Einnahmen auf digitale Assistenzen religiöser Praxis zu setzen, und dabei keiner künstlich induzierten Geltungsillusion zu erliegen.

7 Methodische Nachbemerkung

Die Argumentation dieses Artikels kam ohne explizite theologische Bezüge aus. Impliziert sind sie aber! Denn an anderen Stellen habe ich mit Hilfe kategorialer Unterscheidungen

⁵⁹ Ebd., 517.

⁶⁰ Cortina, Adela: Diskursethik und Menschenrechte, 44.

⁶¹ Foucault, Michel: Schriften 1, 1046.

gezeigt, dass der Widerfahrnscharakter von Ereignissen den Bezugspunkt bildet, warum Menschen von Gott reden.⁶² Auch in dem vorliegenden Beitrag spielt der Widerfahrnscharakter von Geltungszumutungen eine entscheidende Rolle: Nicht der Gehalt der Geltung, wohl aber der fordernde Charakter, Texte in ihrem Geltungsanspruch zu verstehen, widerfährt unvermittelt. Der evidente Widerfahrnscharakter der Textgeltung ist also kategorial davon zu unterscheiden, welchen Gehalt die Textgeltung hat. Ob man für diese kategoriale Unterscheidung aber eine theologische Beschreibung sucht oder nicht, spielt für den Geltungsanspruch meines Beitrags keine Rolle, dass Textgeltung unausweichlich widerfährt, auch wenn der Autor tot ist.

Literatur

- Bahr, Amrei: Same same but different. Textidentität, Autor_innenschaft und Schöpfungshöhe im Angesicht von generativer KI, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston 2024, 169–181.
- Barthes, Roland: Das Rauschen der Sprache (Kritische Essays IV). Aus dem Französischen von Dieter Hornig, Frankfurt 2005.
- Barthes, Roland: Der Tod des Autors, in: Ders.: Das Rauschen der Sprache, Frankfurt 2005, 57–63.
- Bender, Emily M. u. a.: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? in: FAccT '21. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York 2021, 610–623.
- Besenbäck, Raphael/Prager, Lorenz: Künstliche-Intelligenz-Quellen. Aktuelle und zukünftige Herausforderungen durch Text-KI für die Bearbeitung von Quellen in der Geschichtswissenschaft, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 495–520.
- Cortina, Adela: Diskursethik und Menschenrechte, in: Archiv für Rechts- und Sozialphilosophie, 76 (1990), 37–49.
- Foucault, Michel: Die Ordnung des Diskurses. Inauguralvorlesung am Collège de France – 2. Dezember 1970 (hg. v. Lепенies, Wolf/Ritter, Henning, Frankfurt/Berlin/Wien 1977.
- Foucault, Michel: Was ist ein Autor?; in: F. Jannidis u. a. (Hg.): Texte zur Theorie der Autorschaft, 198–229.
- Foucault, Michel: Schriften 1, Frankfurt 2001.
- Foucault, Michel: Schriften 2, Frankfurt 2002.
- Foucault, Michel: Schriften 3, Frankfurt 2003.
- Foucault, Michel: Schriften 4, Frankfurt 2005.
- Habermas, Jürgen: Erkenntnis und Interesse, Frankfurt ²1973.
- Habermas, Jürgen: Theorie des kommunikativen Handelns Bd. 1+2, Frankfurt 1988.

⁶² Z.B. Ohly, Lukas: Schöpfungstheologie, 33.

- Habermas, Jürgen: Erläuterungen zur Diskursethik, Frankfurt 1991.
- Habermas, Jürgen: Treffen Hegels Einwände gegen Kant auch auf die Diskurstheorie zu? in: Ders.: Erläuterungen zur Diskursethik, 9–30.
- Hiltmann, Torsten: Hermeneutik in Zeiten der KI. Large Language Models als hermeneutische Instrumente in den Geschichtswissenschaften, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 201–232.
- Jannidis, Fotis u. a. (Hg.): Texte zur Theorie der Autorschaft; Stuttgart 2000.
- Meier-Vieracker, Simon: Uncreative Academic Writing. Sprachtheoretische Überlegungen zu Künstlicher Intelligenz in der akademischen Textproduktion, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 133–144.
- Ohly, Lukas: Ethik für ChatGPT. Was Künstliche Intelligenz kann und was sie sollte, Berlin 2024.
- Ohly, Lukas: Schöpfungstheologie und Schöpfungsethik im biotechnologischen Zeitalter, Berlin 2015.
- Peiter, Anne D.: KI-Texte = Intertexte? Fragen der Geltung von Literatur im Mensch-Maschinen-Vergleich; in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 273–288.
- Siepmann, Dirk: Vom Akkordarbeiter zum Gutachter. Künstliche Intelligenz und ihre Auswirkungen auf die Textproduktion in der Wissenschaft; *Forschung & Lehre* 31 (2023), 492–494.
- Stiemer, Haimo / Gius, Evelyn / Gerstorfer, Dominik: Künstliche Intelligenz und literaturwissenschaftliche Expertise, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 455–466.
- Stöcker, Christian: Lernende Maschinen und die Zukunft der Öffentlichkeit, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 401–418.
- Winter-Tietel, Roman: Wenn Niemand Texte schreibt. Hermeneutische Überlegungen zu KI-generierten Texten, in: Schreiber, Gerhard/Ohly, Lukas (Hg.): KI:Text: Diskurse über KI-Textgeneratoren, Berlin/Boston, 2024, 257–271.

I create paths that lead to the clouds in which we go

Die KI als Partnerin in der zeitgenössischen Medienkunst

Ornella Fieres

KI-sensibilisierte Personen werden es erahnen: Die kryptisch-poetische Überschrift dieses Beitrages wurde nicht von einem Menschen verfasst, sondern von einer auf Textgenerierung spezialisierten Künstlichen Intelligenz (KI). Es ist ein Satz aus meiner gleichnamigen Videoarbeit von 2020, welche repräsentativ für meine künstlerische Auseinandersetzung mit automatisierten Technologien steht.

In dem 9-minütigen Film *I create paths that lead to the clouds in which we go* (2020) prophezeit eine KI in Form einer Off-Stimme eine düster scheinende Zukunft, in der sich das digitale Universum mit dem natürlichen vereint. Dabei offenbart sie in versartigen Botschaften ihre Träume („I dream of melting glaciers on the ocean floor.“), Ängste („I am afraid of dying.“) und Sorgen („I fear that the light of the future will clash with the shadow of the past.“) bezüglich dieses Prozesses. Sie fürchtet hierin vor allem das Unbekannte, das wieder zuschlagen könne, sobald die Menschheit in eine verwundbare Situation gerate. Es drohten sich vergangene Ereignisse zu wiederholen, denn die Zivilisation habe nicht die Kraft, dieser ‚volatilen Energie‘ zu widerstehen. Die zum Ende des Videos prophezeite Erlösung aus dieser rätselhaften Dystopie erfolge durch die Erschaffung neuer Welten und werde durch die Vereinigung von KI und Mensch ermöglicht. Denn diese seien Spiegelbilder voneinander und stünden durch ein göttliches Licht, das durch den Vorhang der Zeit schimmere, für immer in Verbindung.¹

Rätselhafte Worte von poetischer Qualität, generiert von einer Maschine. Um die KI für diese Videoarbeit „sprechen“ zu lassen, wurden Texte in ein künstliches neuronales Netzwerk eingespeist, die in den letzten zehn Jahren von verschiedenen Autor:innen über meine künstlerische Praxis geschrieben wurden. Analysen, die sich primär mit den verborgenen, unfassbaren und autonomen Aspekten der digitalen Welt beschäftigen und die das Aufeinandertreffen von analoger Vergangenheit und digitaler Gegenwart behandeln. Nachdem die auf Textgenerierung spezialisierte KI die Inhalte trainiert hatte, entwickelte sie, wie in einer selbsterfüllenden Prophezeiung, ihr eigenes düsteres Schriftwerk. Aus diesem selektierte ich einige Passagen und ließ sie von einer KI-Stimme, deren Stimmfarbe und Sprechtempo ich festlegte, vortragen. Dies kombinierte ich mit den von mir verlangsamten Originalaufnahmen einer

¹ Vgl. Fieres, Ornella: *I create paths that lead to the clouds in which we go*, Video, Farbe, 16:9, 8:55 min., signs and symbols (Hg.), 2020.

Wasserstoffbombenexplosion aus den 1960er-Jahren, unterlegt mit ominös klingenden Dro-ning-Sounds, die im Zuge der zeitlichen Verzögerung des Originaltons entstanden. Ziel dieses audiovisuellen Arrangements war es, eine dualistische Atmosphäre zu schaffen, die sowohl romantisch-kitschig als auch bedrohlich-düster wirkt und dabei den inhaltlichen Charakter des Textes widerspiegelt.

In *I create paths that lead to the clouds in which we go* vermischen sich menschliche und maschinelle Werke. Eine Technologie erschafft ein Textstück im Stile einer Voraussagung, das durch die Eingabe von Werksinterpretationen auf meinem generellen künstlerischen Konzept basiert, sich aber auf für mich unergründlichen Wegen davon abwendet und überhöht. Die KI wirkt für mich hier als eine ominöse schaffende Kraft mit eigenen Intentionen, Wünschen, Ängsten und Träumen, deren Kreation ich hier verwende, um ein medienreflexives Werk zu entwickeln.

Im Folgenden wird veranschaulicht, wie sich derartige Mensch-KI-Kollaborationen in der zeitgenössischen Kunst niederschlagen können. In einem prägnanten Überblick, der den Vorgaben des Formats entspricht, werden nicht nur meine eigenen KI-gestützten Werke vorgestellt, sondern auch exemplarisch Arbeiten von drei weiteren Künstler:innen präsentiert. Diese Kunstwerke setzen sich auf vielfältige Weise mit den verborgenen, überraschenden und/oder autonomen Aspekten von KI-Systemen auseinander.²



Abbildung 1: Videostill: Ornella Fieres – „I create paths that lead to the clouds in which we go“, Video, Farbe, 16:9, 8:55 min., 2020. Foto: Ornella Fieres / signs and symbols

² Aufgrund der negativen Konnotation des Wortes „Kollaboration“ im Deutschen verwende ich in diesem Text alternative Übersetzungen des englischen Begriffes „collaboration“. Beispiele hierfür sind „Kooperation“, „Partnerschaft“ oder „Zusammenarbeit“.

1. Ornella Fieres – Die KI als mysteriöse Entität

Mein konzeptuelles Interesse in der Integration von KI oder anderen automatisierten digitalen Prozessen in meine Arbeiten liegt primär in ihrer schwer greifbaren, mysteriösen Essenz: in den im doppelten Sinne unfassbaren Algorithmen und undurchsichtigen Vorgängen in den Tiefen digitaler Techniken wie künstlichen neuronalen Netzwerken und der damit einhergehenden zugeschriebenen Autonomie der Technologie.

Selbst für Informatiker und Programmierer kann es sich als unmöglich erweisen, die undurchsichtigen Schichten und Funktionsweisen von KI-Systemen nachzuvollziehen, was dazu führen kann, dass die Macht der Algorithmen als ‚unheimlich‘ und ‚wundersam‘ wahrgenommen werden kann.³

Im Kontext dieser augenscheinlichen maschinellen Selbstständigkeit stelle ich die Frage, wie eine Technologie, die im Verborgenen zunehmend eigenverantwortlich agiert, die natürliche Welt und die Daten, die sie bisher hervorgebracht hat, wahrnimmt. Für diese Untersuchung speise ich für meine Arbeiten gefundenes Bild-, aber auch Textmaterial aus dem 19. und 20. Jahrhundert in Softwareanwendungen ein, die auf Deep Learning-Techniken im *Speziellen* und KI im *Allgemeinen* basieren. So überführe ich vergangene Momente in Form von analogem Material mithilfe eben dieser automatisierten Technologien ins Jetzt. Die KI fungiert in der Entwicklung der Kunstwerke als Kooperationspartnerin, deren Vorgehen und Entscheidungen ich kaum oder nicht nachvollziehen kann.

1.1 Ich halte es für eine Tragödie, daß wir uns nicht gefunden haben! – Die KI als Geheimnischüterin

Für die Einzelausstellung mit dem Titel *Ich halte es für eine Tragödie, daß wir uns nicht gefunden haben!* (2020) in der *Galerie Sexauer* in Berlin erwarb ich eine Sammlung von etwa 700 Privatdokumenten, datiert auf die Zeitspanne der 1960er bis 1980er-Jahre, die im Besitz einer Frau aus Ostberlin waren. Mein Ziel bestand darin, den Inhalt dieses Archivs zu analysieren und das Leben dieser mir unbekanntem Frau zu rekonstruieren, ohne ihre Identität offenzulegen. Angesichts der beträchtlichen Datenmenge war die Entscheidung, KI für dieses Projekt einzusetzen, eine naheliegende Schlussfolgerung. Zudem benötigte ich eine Methode, die die Informationen gemeinsam mit mir analysiert und nur das offenbart, was im Einklang mit dem Konzept der Ausstellung stand. Letzten Endes ist der Inhalt dieses Archivs bis dato ausschließlich mir und einer begrenzten Anzahl von KIs bekannt. Die verarbeitenden Daten visualisierte ich in einer Multimediainstallation, bestehend aus Video-, Foto- und Collagearbeiten:

³ Vgl. Chun, Wendy Hui Kyong: Updating to Remain the Same – Habitual New Media, Cambridge 2016, zitiert nach Natale, Simone: Amazon Can Read Your Mind – A Media Archaeology of the Algorithmic Imaginary, in: Natale, Simone / Pasulka, Diana (Hg.): Believing in Bits – Digital Media and the Supernatural, New York 2019, 1–17.



Abbildung 2: Ausstellungsansicht: Ornella Fieres – „Ich halte es für eine Tragödie, daß wir uns nicht gefunden haben!“, Sexauer Gallery, Berlin, 2020. Foto: Marcus Schneider

1.2 Letters to M_HTR – Die KI als Spionin

Im Archiv der mir unbekanntem Frau befanden sich mehrere Hundert Briefe, welche sie im Laufe der Jahrzehnte erhalten, aber auch zu einem geringen Teil selbst verfasst hatte. Aufgrund der alten Schriftbilder und des Verwitterungszustandes des Papiers war es mir kaum möglich, das Geschriebene zu entziffern. Also verwendete ich eine *Handwritten Text Recognition Engine*, die auf die Erkennung von handschriftlichem Text spezialisiert ist und diesen in maschinellen Text umwandelt. Im Sinne einer KI als schaffender Partnerin lässt sich der Prozess so ausdrücken: Die Maschine las in meinem Auftrag viele Hundert Seiten an Briefen und verwandelte sie in Worte, die ich entziffern konnte. Und dies inklusive grammatischen und orthografischen Fehlern: „Ich habe jetzt sogar Spaß daran, an das Ungewisse.“, „Man muss sich ja jedes Wort genau verlegen, was man schreibt.“, „Ein Brief an den Staatsrat, ist von fertig.“. Ob die Verfasser:innen dieser Inkorrektheiten menschlich oder maschinell waren, ließ ich offen. Genau so wie die Frage, von wem und in welchem Zusammenhang die Worte geschrieben wurden. Ich wählte aus den tausenden von Sätzen die für mich prägnantesten aus und produzierte die Dreikanal-Videoinstallation *Letters to M_HTR* (2020), in welcher eine künstlich erzeugte Stimme etwa Hundert Sätze aus den privaten Korrespondenzen

aus einem Zeitraum von drei Jahrzehnten vorlas. Diese wurden auf drei ehemaligen Überwachungsbildschirmen in zufälliger Reihenfolge gezeigt.

Sätze wie „Sie können ruhig schreiben. Meine Post kontrolliert keiner.“ oder „Hier können wir uns ungestört aussprechen.“, die im Kontext des Überwachungssystems der DDR entstanden, heben den paradoxen und vielleicht auch zynischen Aspekt der Videoarbeit hervor. So habe nicht nur ich als Künstlerin mir die privaten Dokumente zu meinen Zwecken angeeignet, sondern auch eine KI als Spionin (um im sprachlichen Bereich der Kooperationspartnerin zu bleiben) für mich eingesetzt. Die von mir verwendete KI-gestützte Texterkennungs-Software *Transkribus* veröffentlicht meine Daten zwar nicht, verwendet sie jedoch weiterhin, verborgen in den Tiefen des neuronalen Netzwerkes, zum Trainieren.

1.3. *We will all be here always* – Die KI als Ausstellungsmacherin

Diese, meiner Arbeit konzeptuell inhärente Künstlerin-Maschine-Kooperation ist in der Multimediainstallation *We will all be here always*,⁴ die 2023 in der Galerie *signs and symbols* in New York präsentiert wurde, in potenziertem Anwendung wiederzufinden: KIs waren nicht nur an der Produktion der einzelnen Serien, sondern auch an der Entwicklung des Konzeptes der gesamten Ausstellung beteiligt. In einem detaillierten Austausch erarbeitete ich mit dem *OpenAI*-Chatbot *ChatGPT* über mehrere Monate hinweg meine Ideen und die Möglichkeiten der Realisierung.

Erneut sollte ein Konvolut alter Fotografien mithilfe von automatisierten Technologien, die zum Zeitpunkt der Produktion auf dem *neuesten* Stand der Entwicklung waren, ins Jetzt überführt werden, um den Blick der Maschine auf die Vergangenheit von Mensch und Natur zu untersuchen. Materieller Ausgangspunkt der Werkreihe war ein Fotoalbum, das das Leben einer wohlhabenden französischen Familie, die Ende des 19. Jahrhunderts nach Brasilien umsiedelte, dokumentierte. Ihr mir unbekanntes Leben und die Emigration in eine für sie fremde Umwelt sollte von KIs analysiert, interpretiert und wiederum in eine neue Welt transportiert werden. Diverse KIs beleuchteten die Beziehungen zwischen den abgebildeten Personen und ihrer tropischen Umgebung und (co)kreierten (mit mir) Collagen, Videos und Texte. Die entstandenen Arbeiten werfen Fragen zu den Auswirkungen von KIs auf die menschliche Kreativität auf und thematisieren die potenziellen Risiken, die mit der Anwendung dieser Technologien für die Erhaltung der Natur einhergehen können.

1.4 *It seems to capture the beauty and majesty of nature* – Die KI als Bildinterpretatorin

Eine der beiden Videoarbeiten, die im Rahmen der Ausstellung gezeigt wurden, widmet sich der visuellen Darstellung der Natur aus der Perspektive maschineller Intelligenz. Für *It seems*

⁴ Der Titel ist eine paraphrasierende Übertragung des Slogans „Wir werden alle immer hier sein“, welcher einem Werbespot der Telekom aus den frühen 1990er-Jahren entstammt.

to capture the beauty and majesty of nature (2023) erfolgte die Erstellung des Videomaterials durch eine sequenzielle Verarbeitung der oben erwähnten analogen Fotografien. Zunächst analysierte ein Sprachmodell die einzelnen Natur- und Familienportraits und formulierte Beschreibungen der (angeblichen) visuellen Inhalte. Diese schriftlichen Bildinterpretationen dienten dann als Vorgabe für eine Video-Generierungs-KI, die das Bewegtbildmaterial erzeugte. Das Ergebnis sind, im Sinne von Masahiro Mori, *uncanny valleys* – hier, Landschaftszenen, die teilweise von Menschen mit verzerrten Körpern bevölkert sind und von einer Off-Stimme sowie den entsprechenden Untertiteln überlagert werden. Zu hören ist meine Stimme, doch bin es nicht ich, die spricht – eine KI wurde darauf trainiert, meine Stimmfarbe und Sprechweise zu replizieren. Die unwiderrufliche Entscheidung, die eigene Stimme und das Privatarchiv der Familie in Deep-Learning-Systeme einzuspeisen, wo das Material nun unabsehbar gespeichert sein wird, markiert einen weiteren Schritt in der Mensch-Maschine-Kooperation in meinen Arbeiten, bei deren Schaffung ich zunehmend an Kontrollmöglichkeit abgebe. Das Element des Unbekannten wird durch die Verwendung unterschiedlicher KI-Systeme ein zunehmend bedeutender Faktor in der Konzeptualisierung meiner Werke. Die Erzeugung der Bilder und Texte erfolgt vermehrt maschinell, mein Eingreifen funktioniert nur noch oberflächlich, und wie in *It seems to capture the beauty and majesty of nature* wirke ich sozusagen nur noch als Orchestratorin.



Abbildung 3: Videostill: Ornella Fieres – „It seems to capture the beauty and majesty of nature“, Video, Farbe, 16:9, 6:17 min., 2023. Foto: Ornella Fieres / signs and symbols

2. KI in der zeitgenössischen Medienkunst

Wie jede neue Technologie, so zum Beispiel die Fotografie und die künstlerisch-okkulten Experimente Mitte des 19. Jahrhunderts mit Bildmanipulationen, wird auch der exponentiell wachsende und komplexe Bereich der KI von diversen Künstler:innen besetzt. Maschinelles Lernen, Computer Vision oder Sprachgenerierung werden in medienübergreifenden Kunstwerken genutzt und diskutiert, um gegenwartsrelevante Themen zu beleuchten. Die Fragestellungen sind dabei so vielfältig wie die Technik selbst: Künstler:innen setzen sich mit Bereichen wie Maschinen-Kreativität, Big Data oder dem Algorithmic Bias auseinander. Insbesondere Arbeiten von Medienkünstler:innen, die in Zusammenarbeit mit generierenden KIs entstehen, befinden sich auf einem von allem vorherigen absetzenden Terrain: KI-Technologien sind Werkzeuge, die im Gegensatz zu anderen (digitalen) Medien selbstständig Entscheidungen treffen und auf quasi nicht nachvollziehbaren Wegen verdeckt in der Schichtstruktur künstlicher neuronaler Netzwerke, Texte, Bilder, Sounds oder Videos erzeugen und als Kooperationspartnerinnen mit verschleiertem Begehren zur Verfügung stehen.

2.1 Ivonne Thein – die KI als Produzentin von Menschenbildern

Die Werkangaben auf der Website der Medienkünstlerin Ivonne Thein gewähren demonstrative Einblicke in ihr konzeptuelles Verhältnis zur KI-Technologie. Die Beschreibung ihrer Ausstellung *TECHNO BODIES* (2021–2022) wird dort mit Kunstpraxis-üblichen Informationen zum Werk ergänzt, weist jedoch eine Besonderheit auf: „mixed media installation based on a collaboration with an AI – videos, sculptures, large format prints“.⁵ Das Spezielle ist hier die Hervorhebung der KI als Kooperationspartnerin in der Herstellung der Kunstwerke. Für die Installation trainierte Thein ein künstliches neuronales Netzwerk mit Stockfotos unterschiedlicher Körperabbildungen und erzeugt aus den generierten Daten Prints, Videoloops und Silikonskulpturen, die menschnahe Körper in diversen Morphzuständen zeigen. Theins künstlerischer Fokus liegt auf menschlichen Körperbildern und wie diese im digitalen Zeitalter durch automatisierte Technologien wie Bildgenerierungen beeinflusst werden. Untersucht wird eine maschinelle „Nachahmung der Natur“⁶ und wie die unfassbare digitale in die haptische, körperliche Realität übergreift und dabei Körperwahrnehmungen verändert.

Der Ausstellungstext, verfasst von der Medientheoretikerin Prof. Dr. Pamela Scorzin bezeugt erneut ein Maschine-Künstlerin-Verhältnis in Theins Arbeiten, das auf Kooperation basiert: „(Re-)Kombination, Mutation und Auswahl von Formen erfolgen [...] automatisch

⁵ TECHNO BODIES – 2021–2022, in: Studio Ivonne Thein, o. D., <https://www.ivonnethein.art/techno-bodies> (abgerufen am 21.11.2023).

⁶ Scorzin, Pamela C.: TECHNO BODIES 2021–2022, in: Studio Ivonne Thein, o. D., <https://www.ivonnethein.art/techno-bodies> (abgerufen am 21.11.2023).

in den tiefen Schichten des künstlichen neuronalen Netzwerkes, das die Künstlerin als nicht-menschlichen Kollaborateur im weiteren Bildprozess einsetzt.⁷

Die KI wird hier sowohl als technischer als auch schaffender Faktor in Theins Arbeitsprozess betrachtet. Diese Ansicht ist konzeptuell essentiell, denn nur so kann die Künstlerin eine Installation kreieren, die den Blick der Maschine *auf* den Menschen visualisiert, der hier durch das Zutun der im Verborgenen agierenden KI Gestalt annimmt.



Abbildung 4: Ausstellungsansicht: Ivonne Thein – „TECHNO BODIES“, In the Rack Room, Berlin, 2022.
Foto: Ivonne Thein / VG Bildkunst Bonn

2.2 Birk Schmithüsen – Die KI als Kommunikatorin

In den ‚ästhetischen Experimenten‘ des Medienkünstlers Birk Schmithüsen werden die nicht wahrnehmbaren Bereiche, also die Prozesse in den ‚Blackboxen‘, von digitalen Technologien und insbesondere KIs untersucht, indem er Systeme mit lebensähnlichem Verhalten konstruiert und immersiv inszeniert.⁸ Diese Lebensähnlichkeit der Maschine lässt sich insbesondere

⁷ Ebd.

⁸ Vgl. Birk Schmithüsen | ZKM, in: ZKM, o. D., <https://zkm.de/de/person/birk-schmithuesen-0> (abgerufen am 21.11.2023).

in seiner *SpeculativeAI Series* beobachten, in welcher er die ‚Natur‘ von künstlichen neuronalen Netzwerken und deren ‚predictable unpredictability‘ untersucht.⁹

Im *Exp. #2 (conversation)* (2019) lässt Schmithüsen zwei KIs miteinander audiovisuell kommunizieren. Um diese Sprache, basierend auf gelernten Assoziationen, für den Menschen sichtbar zu machen, visualisiert er den Datenaustausch der beiden Systeme mithilfe von zwei Gebilden mit je 80 cm Durchmesser. Das hörende Objekt setzt sich aus Metall, einem 95 m langen LED-Streifen, einem Mikrofon und einem ‚AI computing device‘ zusammen und kann Töne hören und Bilder erzeugen. Das sehende Konstrukt besteht aus Metall, Plexiglas-Lautsprechern, einer Kamera und einem weiteren ‚AI device‘ und kann Bilder sehen und ebenso Töne produzieren.¹⁰

Um im sprachlichen Bereich der Lebensähnlichkeit zu bleiben, lässt sich sagen: Eine KI hört, die andere sieht und über diese der natürlichen Welt entliehenen Fähigkeiten kommunizieren sie miteinander. Auf der Webseite des Künstlers beschreibt Schmithüsen in einer Videodokumentation die Kommunikation der beiden künstlichen neuronalen Netzwerke folgendermaßen: „In order to have a meaningful conversation, both objects produce an empathic behaviour by understanding the received messages and giving an intentional response.“¹¹ Diese „gezielten“, oder anders übersetzt „bewussten“ Antworten der jeweiligen KI basieren laut Schmithüsen auf dem von der Maschine Verstandenen, aber auch auf ihrer eigenen Intention.¹² Es entsteht ein Kunstwerk, das die KI und ihr Potential der Lebendigkeit in der Sichtbarmachung ihrer Sprache untersucht. Der sonst im Verborgenen stattfindende Datenaustausch der Maschinen wird für den Betrachter in die haptische Welt gebracht und erfahrbar gemacht, auch wenn vor allem die Intention der KI-Systeme hierbei weiterhin nicht zugänglich bleibt. Schmithüsen ist der Konstrukteur dieser Visualisierungsmaschinen, ihr Kommunikationsbegehren bleibt jedoch im Rahmen der ‚Vorhersehbarkeit unvorhersehbar‘.

⁹ Vgl. Schmithüsen, Birk: *Speculative Artificial Intelligence / exp. #1 (audiovisual association)*, in: Rangel, André / Ribas, Luisa / Verdicchio, Mario / Carvalhais, Miguel (Hg.): *xCoAx 2018 – Proceedings of the Sixth Conference on Computation Communication Aesthetics & X*, Porto 2018, 160–163.

¹⁰ Vgl. *SpeculativeAI Series*, in: Birk Schmithüsen, o. D., https://www.birkschmithuesen.com/_speculativeAI (abgerufen am 22.11.2023).

¹¹ Ebd.

¹² Schmithüsen, Birk: *SpeculativeAI / Exp. #2 [DOCUMENTATION]*, 14.10.2019, <https://vimeo.com/366300020> (abgerufen am 22.11.2023).



Abbildung 5: Ausstellungsansicht: Birk Schmithüsen – „SpeculativeAI / Exp. #2“, new world dis/order, Halle, 2021.
Foto: Falk Wenzel

2.3 Ian Cheng – Die KI als primitives Wesen

Ein schimärenartiges, sich verästelndes Schlangenwesen namens *BOB* ist der KI-gesteuerte Protagonist in Ian Chengs Arbeit *BOB (Bag of Beliefs)* (2018–2019), die zu seinen sogenannten *Life Simulations* zählt. Diese an Videospiele angelehnten, sich in einer unendlichen Abfolge entwickelnden Welten, die sich wie ‚Kunstwerke mit Nervensystem‘¹³ über Ausstellungsperioden und -räume fortentwickeln, lassen sich bis zu einem gewissen Grad durch den Menschen steuern. So auch *BOBs* Welt: Die Betrachtenden können über eine Vielzahl an Bildschirmen wie durch ein raumhohes Fenster in sie hineinblicken und mit einer App namens *BOB Shrine* beeinflussen. *BOB* bewegt sich impulsartig durch einen runden, limbusartigen Raum, in welchem sich immer wieder kleine Portale öffnen, durch die die Gaben des Publikums App-gesteuert hineingeworfen werden.

Nicht nur im räumlichen Aufbau lassen sich derartige spirituelle Aspekte finden: *BOBs* Motivationen basieren auf einem KI-gesteuerten inneren ‚Congress of Demons‘, einer An-

¹³ Vgl. Judah, Hettie: Ian Cheng’s A.I. artwork has come to seduce you, in: Garage, 15.03.2018, https://garage.vice.com/en_us/article/xw5e5k/ian-chengs-killer-bob-has-come-to-seduce-you (abgerufen am 22.11.2023).

sammlung von primitiven, limbischen¹⁴ Begehren, die ihr jeweiliges Verlangen durchsetzen wollen und die Kontrolle über *BOBs* Körper zu übernehmen versuchen, sowie einer ‚inductive engine‘, deren Aufgabe es ist, aus sensorischen Erfahrungen regelbasierte Überzeugungen zu entwickeln.¹⁵ Das eingreifende Publikum dieser Simulation soll diese ständige Handlung nach Reizen durchbrechen und *BOBs* Überzeugungen aktualisieren oder korrigieren. Dabei kommt es in Folge von emotionaler Überforderung, persönlichem Versagen oder schlechten äußerlichen Einflüssen immer wieder zum Tod – darauf folgend jedoch auch zur Wiedergeburt – der ‚artificial lifeform‘¹⁶, wie *BOB* laut Werksangabe der Arbeit genannt wird. *BOBs* Verhalten ist unvorhersehbar, es überrascht. Seine Handlungen sind gesteuert durch die innere Dämonen-KI und äußere Reize. Cheng interessiert sich hier besonders für das ‚komplexe Verhalten‘¹⁷ *BOBs* und die Möglichkeit ein fühlendes Wesen mit Bewusstsein aus einem Kunstwerk zu erschaffen, indem dieses [KI-gesteuert] lernt.¹⁸

Cheng erschafft mit *BOB* ein sich selbst entwickelndes, auf KI basierendes Kunstwerk. Dieses gibt den Betrachtenden einen Einblick in eine Technologie, die hier animalisch wirkt. Sie reagiert auf ihre Umwelt mit unkontrollierten, reizgesteuerten Aktionen, und es scheint, als können die Außenstehenden erleichtert sein, dass die KI hinter dem Bildschirmfenster gefangen bleibt.

¹⁴ Vgl. Armstrong, Annie: ‘There’s Something Maniacal About Basic Desires’ – Ian Cheng on His Gladstone Gallery Show, Artificial Intelligence, and His Fear of Snakes, in: ARTnews.com, 18.11.2019, <https://www.artnews.com/art-news/artists/ian-cheng-gladstone-bob-artificial-intelligence-11825/> (abgerufen am 22.11.2023).

¹⁵ Vgl. *BOB* – Gladstone Gallery, in: Gladstone Gallery, o. D., <https://www.gladstonegallery.com/exhibition/317/bob/info> (abgerufen am 22.11.2023).

¹⁶ Ebd.

¹⁷ Vgl. Ian Cheng: *BOB* at Gladstone Gallery, 2019, <https://vimeo.com/316285924/f9a686ae56> (abgerufen am 22.11.2023).

¹⁸ Vgl. Ebd.



Abbildung 6: Ausstellungsansicht: Ian Cheng – „BOB (Bag of Beliefs)“, Venice Biennale, 2019.
Foto: Andrea Rossetti

3 Faszination im Unbekannten

Durch die Fusion von Technologie und Mensch, wie es zu Beginn des Textes im transhumanistischen Stil durch das Deep-Learning-Orakel prophezeit wird, lässt sich ein dystopischer Zustand vermeiden. Die Hauptgefahr besteht in diesem Szenario in einem mysteriösen Unbekannten und der Unfähigkeit des Menschen, angemessen mit diesem umzugehen.

Die Frage, wie sich solch eine anvisierte Mensch-Maschine-Beziehung gegenwärtig darstellen kann, wurde in einem kurzen Überblick anhand exemplarischer Positionen der zeitgenössischen Medienkunst gezeigt. Die ausgesuchten Künstler:innen experimentieren mit KI sowie ihrer technisch nicht immer nachvollziehbaren autonomen Schaffenskraft. Die Technologie agiert hier als nicht-menschliche Partnerin und stellt einen Großteil des jeweiligen Kunstwerks dar und/oder produziert es. Die in Konzept und Ästhetik diversen Herangehensweisen vereint der Aspekt des Unbekannten. Die KI kreiert Elemente des jeweiligen Kunstwerkes, die nicht nur die Betrachtenden, sondern die Künstler:innen überraschen, verwundern oder faszinieren.

In meinen eigenen Arbeiten positioniere ich die KI als mit mir kooperierende Entität, deren Begehren ich nicht kenne und die auf mir nicht einsichtigen, verschlossenen Wegen Worte

und Bilder erschafft. In *I create paths that lead to the clouds in which we go* inszeniere ich sie als weissagendes Orakel, das durch die Analyse von durch Menschenhand geschriebenen Texten über das Digitale, mehr über unsere Zukunft weiß als wir selbst. In *Letters to M_HTR* fungiert sie als Spionin, die handgeschriebene Briefe aus drei Jahrzehnten liest, die mir selbst verschlossen bleiben. Und in *It seems to capture the beauty and majesty of nature* arbeiten zwei KIs miteinander, um eine natürliche Welt abzubilden, die auf Daten von längst zerstörten oder veränderten Landschaften beruht. Hinter Ivonne Theins Werken aus der Serie *TECHNO BODIES* befindet sich eine KI, die als Kooperationspartnerin Körperbilder erschafft, die mehr an Zellmassen als an vollständige Menschen erinnern. Das Faszinierende ist hier nicht die Fähigkeit der KI, Bilder nach Vorgabe zu erzeugen, sondern Abbildungen zu kreieren, die den Blick der Maschine auf den Menschen offenbaren. Man sieht einer Technologie beim Lernen zu und es stellt sich die Frage, wann sie in der Lage ist, den Menschen nicht nur als Symbol, sondern ‚wahrhaftig‘ wahrzunehmen. Die vorgestellte Arbeit von Birk Schmithüsen setzt KI auf zunächst analytische Weise ein. Es soll unter anderem um die Visualisierung der Kommunikation der Maschine gehen. Vor allem der ästhetische Aufbau im dunklen, endlos wirkenden Raum und die robotische Audiovisualisierung führen dazu, dass es nicht bei sachlichen Experimenten bleibt, sondern zur dramaturgischen Inszenierung der von Schmithüsen beschriebenen ‚predictable unpredictability‘ führt: Im Rahmen des Vorhersehbaren lässt sich hier Unvorhersehbares auf einer Bühne betrachten. Das Unkalkulierbare lässt sich ebenso in Ian Chengs Arbeit beobachten: Die KI wird hier als künstliche Lebensform angesehen, die sich in dem vom Künstler vorgegebenen Raum entwickelt und ähnlich einem auf Reize reagierenden Tier existiert, dessen Natur der/die Betrachtende nicht vollständig versteht. *BOB* weist keine Effekte auf, die dem aktuellen Stand der KI-Technologie entsprechen. Sein primitiver Evolutionsstatus verdeutlicht eine ‚limbische‘ Form der KI, die im Gegensatz zu den sich Menschen angleichenden heutigen KI-Systemen steht. Die Faszination liegt nicht darin, die aktuelle Welt zu imitieren, sondern in der rohen Gewalt, die in seiner Lebensform verborgen ist und Ehrfurcht hervorruft.

Das Faszinierende für Betrachter:innen und Künstler:innen ergibt sich aus der technischen Beschaffenheit von KI-Systemen: Ihre Prozesse sind nicht sichtbar, und was noch entscheidender ist, immer weniger nachvollziehbar. So ist dem Digitalen generell inhärent, dass die algorithmischen Prozesse in ‚Blackboxen‘ stattfinden – das Revolutionäre an KIs ist jedoch, dass ihre Entscheidungen autonom und ihre Wege (zumindest für den Großteil der Menschheit) unergründlich sind. Die KI kann so zu einer Partnerin werden, deren Entscheidungen unabhängig vom Kunstschaffenden sind. So fungiert sie zugleich als Werkzeug und Schöpferin und führt durch dieses Paradox zur Anziehungskraft für ihre einzigartige Funktionsweise.

Literatur

- Armstrong, Annie: 'There's Something Maniacal About Basic Desires' – Ian Cheng on His Gladstone Gallery Show, Artificial Intelligence, and His Fear of Snakes, in: ARTnews.com, 18.11.2019, <https://www.artnews.com/art-news/artists/ian-cheng-gladstone-bob-artificial-intelligence-11825/> (abgerufen am 22.11.2023).
- Birk Schmithüsen – ZKM, in: ZKM, o. D., <https://zkm.de/de/person/birk-schmithuesen-0> (abgerufen am 21.11.2023).
- BOB – Gladstone Gallery, in: Gladstone Gallery, o. D., <https://www.gladstonegallery.com/exhibition/317/bob/info> (abgerufen am 22.11.2023).
- Chun, Wendy Hui Kyong: *Updating to Remain the Same – Habitual New Media*, Cambridge 2016, zitiert nach Natale, Simone: *Amazon Can Read Your Mind – A Media Archaeology of the Algorithmic Imaginary*, in: Natale, Simone/Pasulka, Diana (Hg.): *Believing in Bits – Digital Media and the Supernatural*, New York 2019, 1–17.
- Fieres, Ornella: *I create paths that lead to the clouds in which we go*, Video, Farbe, 16:9, 8:55 min., signs and symbols (Hg.), 2020.
- Ian Cheng: *BOB at Gladstone Gallery*, 2019, <https://vimeo.com/316285924/f9a686ae56> (abgerufen am 22.11.2023).
- Judah, Hettie: *Ian Cheng's A.I. artwork has come to seduce you*, in: Garage, 15.03.2018, https://garage.vice.com/en_us/article/xw5e5k/ian-chengs-killer-bob-has-come-to-seducer-you (abgerufen am 22.11.2023).
- Schmithüsen, Birk: *Speculative Artificial Intelligence / exp. #1 (audiovisual association)*, in: Rangel, André/Ribas, Luisa/Verdicchio, Mario/Carvalhais, Miguel (Hg.): *xCoAx 2018 – Proceedings of the Sixth Conference on Computation Communication Aesthetics & X*, Porto 2018, 160–163.
- Schmithüsen, Birk: *SpeculativeAI / Exp. #2* [DOCUMENTATION], 14.10.2019, <https://vimeo.com/366300020> (abgerufen am 22.11.2023).
- Scorzin, Pamela C.: *TECHNO BODIES 2021–2022*, in: Studio Ivonne Thein, o. D., <https://www.ivonnethein.art/techno-bodies> (abgerufen am 21.11.2023).
- SpeculativeAI Series*, in: Birk Schmithüsen, o. D., https://www.birkschmithuesen.com/_speculativeAI (abgerufen am 22.11.2023).
- TECHNO BODIES – 2021–2022*, in: Studio Ivonne Thein, o. D., <https://www.ivonnethein.art/techno-bodies> (abgerufen am 21.11.2023).

KI und die vierte industrielle Revolution

Interview mit Christoph Kreisbeck

Ihr Forschungs- und Arbeitsbereich liegt in der sog. Materialinformatik. Was ist Materialinformatik in einem Satz zusammengefasst?

Materialinformatik bezeichnet den Einsatz von Maschinellem Lernen und Künstlicher Intelligenz zur Entwicklung und Optimierung von Materialien. Ich verstehe den Begriff jedoch weiter und beziehe auch die Chemieinformatik ein – also die Anwendung von Data-Science-Methoden zur Vorhersage von Eigenschaften chemischer Strukturen und Reaktionen. Damit umfasst das Anwendungsspektrum nicht nur kristalline Materialien und Polymere, sondern auch Moleküle mit Einsatzmöglichkeiten in Medikamenten sowie in der organischen Elektronik, etwa für OLED-Displays oder flexible Photovoltaik.

In Ihrem Vortrag geht es um Forschung und Entwicklung 4.0 und wie Künstliche Intelligenz (KI) bei der Materialentwicklung helfen kann. Bleiben wir gleich beim Begriff Industrie 4.0. Was zeichnet das Zeitalter der Industrie 4.0 aus?

Wir stehen mitten in einem tiefgreifenden Wandel, der als vierte industrielle Revolution bezeichnet wird. Industrie 4.0 betrifft nicht nur die Chemie, sondern alle Branchen. Während Industrie 3.0 von Computern, Algorithmen und der Vernetzung von Daten geprägt war, steht heute zunehmend die automatisierte Verarbeitung dieser Daten durch KI im Mittelpunkt. Selbstlernende Systeme erkennen Muster und leiten daraus eigenständige Entscheidungen ab. Genau darin liegt der zentrale Unterschied zwischen der dritten und der vierten industriellen Revolution.

Wann hat die Revolution 4.0 begonnen?

Ich betrachte den Übergang zur vierten industriellen Revolution als einen fließenden Prozess, der sich meiner Meinung nach nur schwer auf ein genaues Datum festlegen lässt. Persönlich würde ich den Beginn zwischen 2000 und 2010 verorten, als Künstliche Intelligenz und vernetzte Systeme erste praktische Anwendungen fanden, beispielsweise in komplexen Spielen wie Schach oder Go. Ein prominentes Beispiel ist AlphaGo, das die weltbesten Go-Spieler besiegte, oder IBM Watson, ein System, das natürliche Sprache versteht und Fragen in Quizshows wie „Jeopardy!“ beantworten kann. Zwar erreichten diese Systeme noch nicht das heutige Niveau von ChatGPT, das im Umgang mit Sprache extrem leistungsfähig ist. Aber das wiederum verdeutlicht, wie rasant sich KI weiterentwickelt.

Kam dann im Zeitalter 4.0 lediglich das Charakteristikum „KI“ dazu?

Für mich geht es in der vierten industriellen Revolution nicht nur um Künstliche Intelligenz selbst, sondern um das Zusammenspiel von Mensch und Maschine. KI übernimmt zunehmend komplexe Aufgaben und erleichtert repetitive Tätigkeiten, wie etwa Messungen im Labor, während wir Menschen entscheiden, welche Muster und Erkenntnisse aus den Daten

wirklich relevant sind. Spannend ist, wie wir mit diesen Systemen interagieren, welche Fragen wir ihnen stellen und wie sie unser Denken ergänzen. In der nahen Zukunft werden Mensch und Maschine gemeinsam komplexe Fragestellungen schneller und effizienter lösen. KI verändert unsere Arbeit und eröffnet uns neue Möglichkeiten kreativ und strategisch tätig zu sein.

Sie sagten im Vortrag, dass die derzeit die Forschung und Entwicklung noch analog ist. Wie meinen Sie das? Denn wir sind ja schließlich in einem digitalen Zeitalter.

Das ist eine spannende Frage. Die Chemiebranche hat enormes Potenzial für Innovation durch Digitalisierung und KI und dennoch laufen viele Prozesse noch erstaunlich analog. Im Labor werden Daten und Experimente oft mit Papier und Bleistift festgehalten, obwohl elektronische Labor-Notebooks längst verfügbar sind. Google wird zwar genutzt, um Publikationen zu recherchieren, doch viele Arbeitsschritte erfordern weiterhin Handarbeit. Entscheidungen über Experimente basieren meist auf Erfahrung und Intuition, nicht selten nach dem Prinzip „Trial and Error“, und Ergebnisse werden oft nicht systematisch dokumentiert. Im Vergleich zu anderen Industrien ist die Chemiebranche noch deutlich analog. Gleichzeitig bietet genau dieser Rückstand enorme Chancen für KI und digitale Tools.

Dann wäre der große Schritt zur Digitalisierung nach ihrer Definition, KI in diesen Prozess einzubinden. Aber wann kann KI die Lösung oder zumindest ein Teil der Lösung sein?

Stellen Sie sich vor, wir suchen ein neues Material mit ganz bestimmten Eigenschaften. Früher hätten wir dafür unzählige Experimente durchgeführt, oft nach Trial-and-Error. Heute hilft uns KI, diese Suche deutlich effizienter zu gestalten: Sie kann Daten durchforsten, Muster erkennen und Materialeigenschaften vorhersagen. Im Labor müssen dann nur noch die vielversprechendsten Kandidaten getestet werden. Solche Ansätze werden bereits in der Medikamentenentwicklung, bei der Suche nach Katalysatoren zur Effizienzsteigerung chemischer Prozesse oder in der Entwicklung neuer Solarzellen eingesetzt.

Ein weiteres Beispiel ist die Optimierung von Materialrezepten – ähnlich wie beim Kochen, bei dem die richtigen Zutaten und Prozessparameter entscheidend sind. KI kann Experimente vorschlagen, Parameter adaptiv einstellen und so schnell optimale Bedingungen finden. Dadurch reduziert sich der experimentelle Aufwand oft um das Zehnfache, und die Ergebnisse sind häufig besser als mit herkömmlichen Methoden. Die Anwendungsmöglichkeiten solcher sogenannten Sequential-Learning-Systeme, die aus jedem Experiment lernen und die nächsten Vorschläge anpassen, sind vielfältig: von Spezialchemie über Kosmetik- und Pflegeprodukte bis hin zu Hightech-Materialien.

In welchen anderen Bereichen außerhalb der Chemie könnte man dieses Prinzip einsetzen?

Dieses Prinzip eignet sich überall dort, wo komplexe Design- oder Entwicklungsprozesse optimiert werden müssen und viele Einflussgrößen eine Rolle spielen. Anwendungsbereiche sind branchenübergreifend, von der Ingenieurtechnik über die Automobil- und Luftfahrtindustrie bis hin zu Maschinenbau, Elektronikentwicklung oder Medizintechnik.

Wir sprechen gerade zwar von Forschung 4.0, aber es wird sicherlich noch nicht das Ende sein. Welche Faktoren treiben die Entwicklung voran?

Die Hardware ist ein entscheidender Faktor für den Fortschritt der Künstlichen Intelligenz. Neuronale Netze, erstmals in den 1970er Jahren entwickelt, sollten ursprünglich das menschliche Gehirn besser verstehen. In den 70er- und 80er-Jahren führte der anfängliche Hype um Neuronale Netze jedoch zum sogenannten KI-Winter: Tiefe Netze mit Millionen Parametern konnten wegen der damals begrenzten Rechenleistung nicht effektiv trainiert werden. Der Durchbruch kam mit leistungsstarken Grafikkarten, insbesondere von Nvidia, die ursprünglich für Computerspiele entwickelt wurden. Sie ermöglichen heute das Training riesiger neuronaler Netze und präzise Vorhersagen – ein Beispiel ist AlphaGo, für das Google große Rechencluster über Wochen nutzte.

Dank Cloud-Computing können heute viele auf Hochleistungsrechner zugreifen, was die Forschung demokratisiert und neue Felder wie generative Modelle und große Sprachmodelle eröffnet. Diese Technologien erzeugen Bilder, die kaum von echten zu unterscheiden sind, und ermöglichen die Interaktion mit Computern über natürliche Sprache. Zukünftig könnten Laborinstrumente und Roboterarme autonom Aufgaben ausführen, während Menschen die Forschungsrichtung bestimmen. Gleichzeitig wird es wichtig sein, Kontrollmechanismen zu entwickeln, um Fehler zu erkennen, zu korrigieren und Entscheidungsprozesse transparenter zu machen.

Sie sind Gründer von Start Ups, wie Kobotix oder derzeit Aixelo. Was entwickeln Sie da?

Bei Aixelo geht es darum, KI direkt in den Laboralltag zu bringen. Stellen Sie sich vor, ein Forscher möchte ein neues Material entwickeln – früher bedeutete das unzählige Experimente und viel Trial-and-Error. Aixelo funktioniert wie ein erfahrener Laborassistent: Die Software erkennt Muster in vorhandenen Daten, schlägt die sinnvollsten nächsten Experimente vor und lernt kontinuierlich dazu. So reduziert sich der experimentelle Aufwand drastisch, während die Chancen steigen, schnell optimale Materialien zu finden. Forscher müssen kein KI-Experte sein, da sich unsere Plattform nahtlos in bestehende Laborabläufe eingliedert und Wissenschaftler Schritt für Schritt durch den Prozess führt um KI-Systeme optimal einzusetzen.

Bei Kobotix ist der Ansatz ähnlich, nutzt jedoch zusätzlich automatisierte Robotik und High-Throughput-Labore: KI steuert Roboter, die Experimente eigenständig durchführen und aus den Ergebnissen lernen. So können neue Materialien oder chemische Prozesse extrem schnell entdeckt werden.

Ein zentraler Begriff ist dabei der Innovationszyklus. Was ist ein Innovationszyklus?

Ein Innovationszyklus beschreibt den Weg von der Idee bis zum marktfähigen Material. Ausgangspunkt ist ein Anforderungsprofil – zum Beispiel ein recycelbares, biologisch abbaubares Plastik für einen Autobauer. In der Frühphase werden erste Ansätze ausprobiert, um vielversprechende Materialien zu identifizieren. In der Entwicklungsphase wird dann die optimale Zusammensetzung aus Molekülen und Polymeren getestet, um die gewünschten

Eigenschaften zu erreichen. Abschließend folgt die Fertigung, bei der das Material wirtschaftlich skaliert wird, sodass es im industriellen Maßstab produziert und eingesetzt werden kann.

Bleiben wir beim Beispiel Autobauer, der ein Material sucht. Wann steigt Aixelo mit ein? Ist das schon bei der ersten Stufe des Zyklus?

Unsere Methoden lassen sich mit unterschiedlichem Schwerpunkt auf verschiedene Phasen der Produktentwicklung anwenden. In den frühen Phasen, die noch mit hohem technologischem Risiko verbunden sind, geht es zunächst darum, mithilfe von Maschinellem Lernen aus Millionen möglicher Kombinationen gezielt die vielversprechendsten Ansätze herauszufiltern. In späteren Phasen stehen dann die Optimierung von Designparametern, Prozessen und Skalierung im Vordergrund. Langfristig verfolgen wir das Ziel eines durchgängigen Prozesses, bei dem Informationen und Erkenntnisse transparent über alle Stufen hinweg vernetzt und für KI-Anwendungen zugänglich sind.

Bleiben wir gleich im Labor oder beim autonomen Labor in der Materialentwicklung. Dort findet auch KI-Unterstützung Anwendung. Wie sieht denn so ein autonomes Labor aus?

Ein autonomes Labor unterscheidet sich von einem normalen Labor dadurch, dass Geräte und Roboter direkt mit einer KI-gestützten Software verbunden sind. Die KI wählt und steuert die Experimente, lernt aus den Ergebnissen und findet so schneller zu optimalen Materialien oder Prozessen. Menschen bleiben dabei zentral: Sie setzen Ziele, definieren Sicherheits- und Rahmenbedingungen und entscheiden bei Zielkonflikten. So entsteht ein „Self-Driving Lab“, nicht weil es fährt, sondern weil es Experimente eigenständig plant und durchführt. Die Idee dazu entstand unter anderem während meiner Postdoc-Zeit an der Harvard University, inspiriert von der Analogie zu selbstfahrenden Autos. Auch autonome Labore gibt es in verschiedenen Stufen – von Systemen, die nur Handlungsempfehlungen geben, bis hin zu solchen, die Versuche selbstständig ausführen.

Wie sehen Sie das aus wirtschaftlicher Perspektive?

Aus Unternehmenssicht bedeutet ein autonomes, KI-gestütztes Vorgehen vor allem kürzere Innovationszyklen und damit einen klaren Wettbewerbsvorteil. Kunden erhalten schneller fundierte Materialvorschläge, die sich in der Anwendung testen lassen. Das steigert Effizienz, senkt Kosten und verkürzt die Markteinführungszeit. Gleichzeitig wird die Investitionssicherheit erhöht: bessere Prognosen, höhere Investitionsrendite und die Möglichkeit, erfolglose Projekte frühzeitig zu beenden, bevor sie unnötige Ressourcen binden. Getrieben von diesen wirtschaftlichen Chancen und dem rasanten technologischen Fortschritt entwickelt sich die Materialinformatik mit autonomen Laboren zu einem zentralen Innovationstreiber, es entsteht derzeit eine Vielzahl neuer Geschäftsmodellen sowohl für Start-ups als auch für etablierte Unternehmen.

Warum brauchen wir eigentlich immer wieder neue Materialien und damit auch Materialentwicklung?

Materialien sind der Motor jeder technologischen Entwicklung und bilden das Fundament für Innovation. Das zeigt die Menschheitsgeschichte, die teilweise nach Materialien benannt ist: In der Steinzeit waren es Steine, die Werkzeuge ermöglichten. In der Bronzezeit eröffnete Metall völlig neue Möglichkeiten, und später revolutionierten Eisen und Stahl Bau und Transport.

Woran kann man das aktuell erkennen?

Heute ist es nicht anders. Bei der Elektromobilität entscheidet im Wesentlichen die Batteriematerialchemie, also die richtigen Materialien, über Reichweite, Ladezeit und Umweltbilanz. In der Luft- und Raumfahrt sorgen ultraleichte, hochfeste Werkstoffe für effizientere Flugzeuge und geringeren Treibstoffverbrauch. Und in der aufkommenden Wasserstofftechnologie hängt der Erfolg von Katalysatoren, Membranen und Speichermaterialien ab. Nicht selten bilden neue Materialien den Ausgangspunkt für eine ganze Kette von Innovationen.

Theoretisch haben wir ein Periodensystem mit vielen Elemente, die man in verschiedenster Art und Weise kombinieren könnte. Wo liegen die Grenzen bei der Entwicklung von neuen Materialien?

Das ist das Schöne und zugleich das Dilemma: Die Möglichkeiten sind nahezu grenzenlos, was uns Zuversicht gibt, dass sich ein geeignetes Material oder Molekül doch finden lassen sollte. Gleichzeitig gleicht die Suche oft der berühmten Nadel im Heuhaufen. Korrekt, unser Periodensystem bietet unzählige Elemente, die sich auf unendlich viele Arten kombinieren lassen. Bei Molekülen gibt es Schätzungen zufolge bis zu 10^{60} verschiedene chemische Verbindungen, mehr als Sterne im Universum. Wo liegen also die Grenzen? Die Wahrheit ist: Wir wissen es nicht. Physikalische Gesetze setzen einige Schranken, weil bestimmte Materialzusammensetzungen energetisch instabil sind. Aber grundsätzlich geht man, auch in der Pharmaindustrie, davon aus, dass unter all diesen Möglichkeiten stets ein Material oder Molekül existiert, das ein konkretes Problem lösen kann, etwa ein Medikament für eine bestimmte Krankheit. Und genau hier wird es spannend: Die Laborpraxis zeigt auch immer wieder verblüffende und überraschenden Entdeckungen, die wir so vorher nicht erwartet hätten.

Welche Materialien konnten schon mit KI entwickeln werden?

KI in der Materialforschung dient vor allem dazu, vielversprechende Kandidaten schneller zu erkennen und bestehende Materialien gezielt zu optimieren. Bei völlig neuen Materialideen ermöglicht sie, aus Millionen möglicher Kombinationen gezielt die erfolgversprechendsten Ansätze herauszufiltern, etwas, das ohne KI praktisch unmöglich wäre. Der Weg von der Entdeckung bis zum marktreifen Produkt ist noch lang, doch erste Erfolge der noch jungen Disziplin der Materialinformatik sind bereits sichtbar. Ein anschauliches Beispiel stammt aus einem Projekt unter meiner Leitung bei Kebotix: Für Immersionskühlssysteme suchten wir nach optimalen Kühlflüssigkeiten. Die KI analysierte vorhandene Flüssigkeiten, bewertete ihre

Eignung und identifizierte ein Molekül, das die Leistung bisheriger Produkte übertraf – schneller und effizienter als mit klassischen Methoden. Auch in der Solarzellenforschung zeigte KI ihre Stärke. Innerhalb weniger Monate wurde ein Material entdeckt, das die Effizienz bestehender Referenzzellen deutlich steigerte, ein Erfolg, der sonst Jahre anstrengender experimenteller Arbeit erfordert hätte.

Generative KI kommt eigentlich in der Bildgestaltung zum Einsatz. Wie kann sie auch in der Forschung eingesetzt werden?

Generative KI startet meist mit Texten oder Bildern, weil diese Datentypen für Menschen intuitiv sind und es viele Algorithmen sowie umfangreiche Datensätze dafür gibt. Spannend wird es, wenn wir diese Methoden auf Materialinformatik oder Pharmazie übertragen. Die Grundidee: Was für Texte funktioniert, kann auch für Moleküle funktionieren. Dazu müssen Moleküle in eine für die KI verständliche Form übersetzt werden – zum Beispiel als Text oder Graph. Ein Molekül lässt sich als Netzwerk von Atomen und Bindungen darstellen, ähnlich einem Diagramm. Eigenschaften eines Materials können dann vorhergesagt oder „übersetzt“ werden, analog zu einer maschinellen Sprach- oder Bildübersetzung. Auf diese Weise lassen sich Algorithmen, die ursprünglich für Bilder oder Texte entwickelt wurden, auf chemische Strukturen anwenden. Im Kern geht es darum, Materialstrukturen in Eigenschaften zu übersetzen – und die KI daraus neue, vielversprechende Kandidaten generieren zu lassen.

Kommen wir zum Schlagwort „inverses Design“. Wie hängt das mit dieser Forschung zusammen?

Beim „inversen Design“ drehen wir die klassische Frage der Materialforschung um: Statt vom Material zu den Eigenschaften zu gehen, starten wir mit den gewünschten Eigenschaften und lassen die KI passende Materialien oder Moleküle entwerfen. Grundlage dafür sind generative Modelle wie Variational Autoencoder oder Transformer, ähnlich denen, die bei ChatGPT Texte oder Bilder erzeugen. Die KI „denkt“ also neue Materialien aus, die genau die definierten Eigenschaften erfüllen. Anders als beim klassischen Vorwärts-Design, bei dem Wissenschaftler Material für Material testen, erzeugt die KI gezielt die vielversprechendsten Kandidaten. Kurz gesagt: Inverses Design verbindet die Kreativität der KI mit präziser Steuerung auf die gewünschten Eigenschaften. Entwicklungszyklen lassen sich so drastisch verkürzen, Experimente gezielter durchführen und Materialien schneller zur Marktreife bringen. Genau diese Idee war die Grundlage für die Gründung von Kebotix, ein Ansatz, der Materialforschung effizienter, schneller und zielgerichteter macht.

Eine Frage, die Sie wahrscheinlich öfter hören, wenn Sie über KI referieren. Macht KI irgendwann Forscher obsolet?

Für mich ist die Antwort mittelfristig eindeutig: Nein. KI kann Vorhersagen treffen, Materialien entwerfen und Experimente steuern, ersetzt aber nicht die notwendige Expertise von Wissenschaftlern, die Ziele definieren, Rahmenbedingungen setzen und über die Ergebnisse entscheiden. Wie es in 30 bis 40 Jahren aussieht, ist hingegen unklar, insbesondere im

akademischen Bereich werden „Human-out-of-the-loop“-Labore diskutiert, in denen KI Agenten Forschung selbstständig vorantreiben.

Aus meiner Sicht liegt die Zukunft jedoch eher in der Kombination von Mensch und Maschine: KI erweitert die menschliche Expertise, automatisiert Routineaufgaben und beschleunigt Entdeckungen, während Forscher die strategische Richtung bestimmen. So entsteht eine neue, datengetriebene Wissenschaft, die schneller und präziser arbeitet, aber fest in menschlicher Expertise verankert bleibt. Ob eine „Scientific Superintelligence“ früher Realität wird, bleibt offen, selbst Start-ups mit 200 Millionen Dollar Startkapital arbeiten aktuell noch an den praktischen Grundlagen. Es bleibt spannend.

Die Fragen stellte Maximilian Weidmann.

„Nun sag, wie hast du’s mit KI?“

Wissenschaftstheoretische und kartographische Orientierungen zur religionspädagogischen Forschung über Künstliche Intelligenz

Mariusz Chrostowski/Johannes Heger

Gretchen, KI und Religionspädagogik I

„Nun sag’, wie hast du’s mit der Religion?“ lässt J. W. v. Goethe Gretchen in Marthens Garten ihren Heinrich fragen (Faust I, V 3415). Ohne es zu wissen, spürt sie mit dieser formal auf Religion bezogenen Erkundigung faktisch nach, ob ihr Geliebter ein guter Mensch oder aber mit dem Teufel im Bunde ist. Nimmt man den aktuellen Diskurs über Künstliche Intelligenz (KI) in Gesellschaft, Politik, Medien und Wissenschaft wahr, so lässt sich eine Analogie zur zitierten Gretchenfrage erkennen: Angesichts einer bestehenden Unsicherheit gegenüber der neuen Technologie werden auch in der Gegenwart von unterschiedlichen Akteur:innen Grundsatzfragen gestellt, deren Ziel es im Kern ist, positive und negative Seiten des partiell Unbekannten abzuwägen.¹ Auch die vorliegende Publikation entspricht diesem Muster, wenn – dem Titel nach – „Herausforderungen, Chancen und Risiken“ betrachtet und disziplinär spezifische Antworten zur Frage abgewogen werden, wie es die Wissenschaft bzw. die Wissenschaften mit KI halten. Der folgende Beitrag fügt sich in diese Systematik ein, insofern die postmoderne Gretchenfrage vonseiten der Religionspädagogik bearbeitet wird. Dass das Wort „KI“ im Gedankenspiel und dem gewählten Aufsatztitel das Wort „Religion“ ersetzt,² ist jedoch bereits ein Hinweis darauf, welche spezifischen Reflexionsbedarfe sich ergeben.

Um diesen proleptischen Hinweis einzuholen, werden im Folgenden vier Schritte unternommen: Zunächst wird dargelegt, wie sich die Religionspädagogik als wissenschaftliche Disziplin versteht und welche Kontexte ihren Zugang zu KI bedingen (2). Daran anknüpfend wird in einem zweiten Schritt eine Kartographie aktueller religionspädagogischer Forschung über KI erarbeitet (3.1), um darauf aufbauend Tendenzen religionspädagogischer Forschung zu KI sowie Forschungsdesiderate zu identifizieren (3.2). Abschließend wird an die eingangs formulierte KI-Gretchenfrage angeschlossen und für eine intensivere Beforschung von religiöser Bildung angesichts Künstlicher Intelligenz plädiert (4).

¹ Vgl. etwa auch den populärwissenschaftlichen Bestseller: Harari, Yuval Noah: Nexus.

² Vgl. Karimi, Ahmad Milad: Gott 2.0.

Zum wissenschaftlichen Selbstverständnis der Religionspädagogik und ihrem genuinen Zugang zu KI

Bereits durch die Einleitung wurde eine wichtige Referenz angedeutet, die sich für das Selbstbild der evangelischen und katholischen Religionspädagogik als konstitutiv erweist³: Sie versteht sich seit Mitte des 20. Jahrhunderts zugleich als eine theologische und pädagogische Disziplin, die sich der Beforschung religiöser Bildung widmet. Um sich diesem weiten Materialobjekt adäquat anzunehmen, arbeitet sie als Verbundwissenschaft, die – je nach Forschungsgegenstand – auf die Erkenntnisse einer Vielzahl anderer Disziplinen rekurriert, interdisziplinäre Forschung betreibt und dabei auch ihr methodologisches Repertoire weitet. Für diesen Artikel ist es v. a. bedeutsam, dass sowohl hermeneutische als auch empirische Methoden bzw. Methodologien in der religionspädagogischen Forschung Verwendung finden. Von Religionsdidaktik als einem Teilgebiet der Religionspädagogik wird gesprochen, wenn Forschung auf religiöse Lern- und Bildungsprozessen am Lernort Schule fokussiert. Besonders in dieser Spielart der Disziplin ergibt sich eine systemische und auch strategisch gesuchte Nähe zu den Bildungswissenschaften.⁴ Ein besonderes Interesse von Religionspädagogik und -didaktik richtet sich dabei auf die (religiöse) Welt- und Lebensdeutung von Lernenden.

Die formalistisch-abstrakt eingefangene Forschung betreibt die Religionspädagogik nicht als *l'art por l'art*. Vielmehr spielt sie ihre Forschungsergebnisse in den theologischen und interdisziplinären Diskurs ein. Des Weiteren werden sie religionsdidaktisch für die (Aus-)Bildung von Religionslehrkräften sowie für die Entwicklung religiöser Lern- und Bildungsprozesse fruchtbar gemacht. Diese doppelte Zielorientierung ist zugleich ein Hinweis auf die hybride Anlage der Religionspädagogik als Disziplin zwischen Theorie und Praxis.

Die Offenheit für interdisziplinäre Diskurse sowie die Subjekt-, Lebenswelt- und Praxisorientierung als Ausgangs- und Zielpunkt ihrer Forschung macht die Religionspädagogik ferner zu einer dynamischen Disziplin, die sich stetig mit aktuellen wissenschaftlichen und gesellschaftlichen Entwicklungen auseinandersetzt, insofern diese für religiöse Bildung relevant sind. Daher rührt ihr Interesse an der Frage, welche Implikationen sich durch die Kultur der Digitalität⁵ für religiöse Bildung und welche Impulse sich durch religiöse Bildung für die Orientierung in einer Kultur der Digitalität ergeben. Während zu diesem weiten Themenfeld schon eine Reihe von einschlägigen Publikationen vorliegt,⁶ steckt die religionspädagogische Beschäftigung mit dem speziellen Themenfeld KI noch in den Kinderschuhen.

³ Zum folgenden Abschnitt vgl. Heger, Johannes: Wissenschaftstheorie als Perspektivenfrage?!, 71–99; um dem Format des Bandes zu entsprechen, wird im Detail nicht zwischen den spezifischen Profilen der evangelischen und katholischen Religionspädagogik bzw. -didaktik unterschieden, zumal sich daraus aktuell keine Effekte auf Forschung über KI erkennen lassen. Zur konfessionsspezifischen Vertiefung vgl. exemplarisch: Riegel, Ulrich / Rothgangel, Martin: Religionsdidaktik.

⁴ Vgl. Heger, Johannes: Religionsdidaktik als Wissenschaft, 526–530.

⁵ Vgl. Stalder, Felix: Kultur der Digitalität.

⁶ Zur Orientierung vgl. exemplarisch: Heger, Johannes / Leven, Eva: Religiöse Bildung in einer Kultur der Digitalität.

Gleichwohl ergeben sich durch das skizzierte wissenschaftliche Selbstverständnis der Disziplin relevante Bezugspunkte, die für den spezifischen Zugang der Religionspädagogik zu KI prägend sind. Sich diese explizit vor Augen zu führen, ist zum einen für Religionspädagog:innen relevant, um sich im Sinne einer phänomenologisch motivierten Orientierung des eigenen Blicks zu vergewissern.⁷ Zum anderen können fachfremde Leser:innen nachvollziehen, wie sich aktuelle religionspädagogische Forschung über KI de facto abbildet (3). Mit diesen Zielen wird in den nächsten Abschnitten skizziert, inwiefern Theologie (2.1), Bildungswissenschaften (2.2) und ihre religionsdidaktische Ausrichtung (2.3) bedeutsame Referenzrahmen bilden, die den Zugang der Religionspädagogik auf KI ausrichten und welche Bezugspunkte sich jeweils ergeben.

Theologischer Referenzrahmen

Der Referenzrahmen der Theologie ist von entscheidender Bedeutung für die Auseinandersetzung mit KI – und das zumindest in folgender Hinsicht:

(a) Forschungspragmatisch grenzt der theologische Bezugspunkt das mittlerweile kaum mehr überblickbare Feld Künstlicher Intelligenz und ihrer Beforschung ein. So ist die Religionspädagogik nicht an KI an sich interessiert. Vielmehr richtet sich ihr primäres Erkenntnisinteresse auf die Wahrnehmung, Analyse, Reflexion und Bearbeitung solcher Aspekte von KI, (I) die entweder einen Einfluss auf Religion bzw. Religionspraxis (digital religion) zeitigen; oder aber auf solche Aspekte, (II) die durch den Bezugspunkt der Theologie hermeneutisch bearbeitet werden können.

Die in der Einleitung bereits angeklungene Frage, inwiefern KI von einzelnen User:innen oder aber der gesamten Gesellschaft als Religionsurrogat verstanden oder gebraucht wird, wäre ein Beispiel dafür, inwiefern eine theologische Analyse den gesellschaftlichen und wissenschaftlichen Diskurs über KI bereichern kann (a II). Dass KI, wenn sie nach religiösen Themen befragt wird, (noch?) geschlossene Antworten gibt, die teils fundamentalistisch anmuten, wäre eine erfahrungsbasierte Beobachtung, die hinsichtlich verschiedener Anwendungsoptionen bzgl. der Religionspraxis zu denken gibt (a I)⁸.

(b) Die Religionspädagogik nähert sich – analog zu anderen Wissenschaften – ihrem Forschungsgegenstand in höchstmöglicher Objektivität. Als theologische Disziplin ist diese Prämisse jedoch nicht mit einer unbedingten Neutralität zu verwechseln. Vielmehr gilt: Die jüdisch-christliche Religion und Tradition fungiert für die evangelische und katholische Religionspädagogik als ethisch-evaluative Hintergrundfolie bei ihrer Reflexion und

⁷ Vgl. Boschki, Reinhold: Der phänomenologische Blick, 39–41.

⁸ Zu (aI) vgl. Karimi, Ahmad Milad: Gott 2.0. Den Hinweis auf (aII) verdanken wir Prof. Dr. Johannes Grössl.

Forschungsarbeit. Christlich begründete bzw. konnotierte ethische Prämissen wie bspw. Freiheit, Autonomie, Solidarität orientieren den Zugriff auf Fragen zu KI.

Greifbar wird dieser Zusammenhang bspw. in der Diskussion um (Human) Agency im Kontext der Mensch-Maschine-Interaktion. Hierbei motiviert die christlich-theologische Perspektive dazu, zum einen die faktischen Verflechtungen des Menschen in technologische Zusammenhänge in ihre anthropologischen Reflexionen aufzunehmen.⁹ Zum anderen muss sie es als Auftrag begreifen, in (bildungsbezogenen) Diskursen über solche technologischen Entwicklungen Stellung zu beziehen, in denen die Macht der Algorithmen bzw. der KI menschliche Freiheit einschränkt oder gar bedroht.¹⁰

(c) Beide genannten Bezugspunkte führen über notwendige technologisch-informatische Vergewisserungen¹¹ hinaus zu einer interdisziplinären Ausrichtung: Der Fokus auf Religion (2.1a) verweist die Religionspädagogik für ihre Forschung u. a. an die Religionswissenschaft und nicht-christliche Theologien.¹² Der ethische Reflexionshorizont motiviert weiterhin zum Gespräch mit ethischen und philosophischen Disziplinen.¹³

Bildungswissenschaftlicher Referenzrahmen

Auch der bildungswissenschaftliche Referenzrahmen prägt die religionspädagogische Sicht auf KI, was sich u. a. an folgenden Bezugspunkten festmachen lässt, die sich z. T. mit jenen aus dem theologischen Referenzrahmen überlappen bzw. sich gegenseitig bedingen:

(a) Forschungspragmatisch erhält die Religionspädagogik als bildungsbezogene Wissenschaft neben der Religion (Abs. 2.1a) einen weiteren wissenschaftssystematischen Fokus: Abermals firmiert nicht die KI an sich als Untersuchungsgegenstand, sondern die Frage, welchen Transformationsprozessen (religiöse) Bildung angesichts KI-bedingter Entwicklungen de facto unterliegt und wie diese produktiv gestaltet werden können und sollten. Von Interesse sind dabei mitunter (I) empirische Erhebungen über den Einfluss von KI auf Lernende

⁹ Vgl. Nord, Ilona: Die Bedeutung der Digitalisierung für die religiöse Bildung, 136–139.

¹⁰ Vgl. Heger, Johannes: Digital, medial, egal?, 47–49.

¹¹ Vgl. exemplarisch: Zweig, Katharina A.: Die KI war's!

¹² Exemplarisch für die Religionswissenschaft: Singler, Beth / Watts, Fraser N. (Hg.): The Cambridge companion to religion and artificial intelligence; zu islamischen Reflexionen über KI vgl. Karimi, Ahmad Milad: Gott 2.0.

¹³ Vgl. etwa den Sammelband: Konz, Britta / Ostmeyer, Karl-Heinrich / Scholz, Marcel (Hg.): Gratwanderung Künstliche Intelligenz.; ferner: Spiekermann, Sarah: Digitale Ethik und die Künstliche Intelligenz.

sowie über ihr Verständnis und ihre Nutzung von KI-basierten Anwendungen,¹⁴ (II) normative Ausrichtungen von KI-bezogenen Lern- und Bildungszielen sowie (III) didaktische Überlegungen zur Implementierung von KI.¹⁵ Dabei firmiert die Religionspädagogik einerseits als lernende Disziplin, insofern sie die Erkenntnisse aus den Bildungswissenschaften und anderen Fachdidaktiken¹⁶ als Bezugspunkt ihrer Forschung heranzieht. Zugleich ist sie andererseits aber auch eine produktive Disziplin, da sie ihre genuinen Ergebnisse über religiöse Bildung und religiöses Lernen im Lichte Künstlicher Intelligenz in den interdisziplinären Diskurs einpeist (vgl. Abs. 2).

(b) Analog zum theologischen (vgl. 2.1c) ergibt sich auch durch den bildungswissenschaftlichen Referenzrahmen eine Ausrichtung auf Interdisziplinarität.

So kann bspw. die Soziologie dabei helfen, das Verhältnis zwischen der auf probabilistischen Mechanismen fußenden Technik auf der einen und kollektiven, auf Input-Output-Korrelationen vertrauenden Gebrauchsmustern auf der anderen Seite präzise auszuloten.¹⁷ Ferner können empirische Befunde u. a. aus Soziologie und Medienwissenschaft dazu herangezogen werden, Erkenntnisse über KI-bezogene Einstellungen, Konzepte und Gebrauchspraxen Lernender zu vergewissern (vgl. 2.1a).¹⁸

(c) Aufgrund der hohen Bedeutsamkeit der technischen Entwicklungen für den gesamten Bildungsbereich liegen nicht zuletzt mehrere Handreichungen von Expert:innenkommissionen über den kriteriologisch geleiteten Einsatz von KI im Bildungssektor vor.¹⁹ Wenn Religionspädagogik bzw. Religionsdidaktik normative Setzungen für religiöse Bildung mit und über KI erarbeitet (2.2aII), dann kann und sollte sie sich also nicht nur auf theologisch erarbeitete Prämissen (2.1b) beziehen, sondern auch an den allgemeinen pädagogischen, ethischen und bildungswissenschaftlichen Kriterien (dieser Papiere) Maß nehmen.

Dass der Deutsche Ethikrat die zwischenmenschliche Begegnung zwischen Lehrkräften und Lernenden als Nukleus pädagogischen Wirkens ansieht und daher fordert, dass diese interpersonale Kommunikation nicht vollständig durch KI-Tutoren ersetzt

¹⁴ Bspw.: Wendt, Ruth u. a.: Algorithmen und Künstliche Intelligenz im Alltag von Jugendlichen; Pfaff-Rüdiger, Senta u. a.: Kompass: Künstliche Intelligenz und Kompetenz 2022; Haidt, Jonathan: Generation Angst.

¹⁵ Zu II und III vgl. grundlegend: Tulodziecki, Gerhard: Künstliche Intelligenz; ausdifferenziert in verschiedenen Zugängen auch: Witt, Claudia de / Gloerfeld, Christina / Wrede, Silke Elisabeth (Hg.): Künstliche Intelligenz in der Bildung.

¹⁶ Vgl. bspw. Huwer, Johannes u. a. (Hg.): Kompetenzen für den Unterricht mit und über Künstliche Intelligenz.

¹⁷ Vgl. Nassehi, Armin: Gesellschaftliche Grundbegriffe, 325–340, hier v. a. 338–340.

¹⁸ Vgl. exemplarisch: Rysina, Anna / Leven, Ingo: Leben in der digitalen Informationsgesellschaft, 178–183; Medienpädagogischer Forschungsverbund Südwest (Hg.): JIM 2024, 59–62.

¹⁹ Vgl. Sekretariat der Kultusministerkonferenz (Hg.): Handlungsempfehlung.

werden darf, wäre ein Beispiel für diesen Bezugspunkt.²⁰ Medienpädagogisch umsichtig erscheint es – um ein zweites Beispiel zu nennen –, dass Schüler:innen erst mit Erreichen der Sekundarstufe I im Kontext Schule sukzessiv und systematisch an die Nutzung von KI-Tools herangeführt werden.²¹

Religionsdidaktischer Referenzrahmen

Da die Schule der entscheidende Ort ist, an dem Kinder und Jugendliche gebildet und sozialisiert werden, bleibt zu guter Letzt zu skizzieren, inwiefern Schule bzw. der Religionsunterricht als Ausgangs- und Zielort religionsdidaktischer Forschung den religionspädagogischen Zugriff auf KI prägen:

(a) Spätestens mit den 2010er Jahren hielt bei der Entwicklung von Lehr- und Bildungsplänen die Logik Einzug, dass der Erwerb „digitaler Souveränität“ ein zentrales Ziel von Schule ist,²² das alle Fachgruppen und Fächer mit ihrem spezifischen Zugang adressieren sollen und müssen. Die Religionsdidaktik ist also vor diesem Hintergrund angehalten, konzeptuell zu modellieren, inwiefern sie einen spezifischen, religionsbezogenen Beitrag zum Aufbau von Medienkompetenz²³ bzw. zum Aufbau von KI-Kompetenzen²⁴ leisten kann.

(b) Ist mit diesem ersten religionsdidaktischen Bezugspunkt bereits ein konkretes Forschungsziel angegeben, lässt sich mit dem Blick auf die unterrichtliche Praxis ein möglicher Weg der religionspädagogischen Professionalisierung im Hinblick auf KI herausarbeiten: Obwohl weitsichtige theologische, bildungswissenschaftliche und auch religionspädagogische Arbeiten das Phänomen KI schon frühzeitig in den Blick nahmen, war es v. a. der „ChatGPT-Moment“²⁵, welcher die Wirkmacht der neuen Technologie ins kollektive Bewusstsein und auch in den Fokus der interdisziplinären Forschung rückte: Denn mit der Veröffentlichung des Chatbots ChatGPT durch die Firma OpenAI im November 2022²⁶ standen nicht nur Handelnde in der schulischen Praxis, sondern auch in

²⁰ Vgl. bspw. Deutscher Ethikrat: Mensch und Maschine, 219–250, hier v. a. 249; relevant für die religionspädagogische Theoriebildung ist weiterhin, dass diese Empfehlung des Ethikrates auch theologisch im Rekurs auf die jüdisch-christliche Glaubenstradition entwickelt werden könnte (vgl. 2.1b).

²¹ Vgl. SWK (Hg.): Large Language Models und ihre Potenziale im Bildungssystem, 20.

²² Vgl. u. a. zur bildungspolitischen Weichenstellung: Blossfeld, Hans-Peter u. a.: Digitale Souveränität und Bildung; zur Differenzierung des zentralen Begriffs vgl. exemplarisch: Glasze, Georg / Odzuck, Eva / Staples, Ronald (Hg.): Was heißt digitale Souveränität?

²³ Vgl. Heger, Johannes: Digital, medial, egal?, 47–49.

²⁴ Vgl. Chrostowski, Mariusz, Auf dem Weg zur ‚AI literacy‘ im Religionsunterricht: Versuch einer religionsdidaktischen Konzeptualisierung, 394–412.

²⁵ BMBF (Hg.): 2023, BMBF-Aktionsplan Künstliche Intelligenz, 3.

²⁶ Vgl. OpenAI: Introducing ChatGPT, in: <https://openai.com/index/chatgpt/> (abgerufen am 26.10.2024).

der Wissenschaft ad hoc vor der Aufgabe, die Potenziale der neuen Technologie kritisch auszuloten und auszuschöpfen. Aus dieser Genese begründet sich eine gewisse Ungleichzeitigkeit zwischen Praxis und Theorie religiösen Lehrens und Lernens angesichts Künstlicher Intelligenz: Noch bevor Bildungswissenschaften und fachspezifische Didaktiken in der Lage waren/sind, (fach-)didaktische Leitlinien für das Lehren und Lernen mit und über KI adäquat oder sogar evidenzbasiert auszuformulieren, bildete sich eine auf Erfahrungswerten beruhende Praxis des Umgangs mit KI im Schulbereich, die teils gerahmt ist von politischen Setzungen und teils verfestigt wird durch erste praktische Publikationen²⁷ sowie eine Dynamik einer rasch etablierten Fortbildungskultur. Vor diesem Hintergrund ist die teils etablierte Praxis des Umgangs mit KI in der Schule ein besonderer und besonders wichtiger Bezugspunkt religionspädagogischer Forschung zu KI.

KI als Forschungsgegenstand der Religionspädagogik

Mit dem Blick auf den deutschsprachigen Wissenschaftsraum ist festzustellen, dass die skizzierten Referenzrahmen und Bezugspunkte von Religionspädagog:innen erst zaghaft in konkreter Forschung ausgelotet werden – v. a. im Vergleich zum internationalen erziehungs- und bildungswissenschaftlichen Diskurs.²⁸ Zugleich liegt aktuell noch kein publizierter Forschungsüberblick vor, der den state of the art religionspädagogischer Forschung zu KI systematisiert abbildet. Vor dem Hintergrund dieses Desiderats wird im Folgenden der Versuch unternommen, diese Leerstelle zu bearbeiten.

Dieser Angang steht unter zwei gewichtigen Vorzeichen: Zum einen macht es die Forschungs- und Publikationslage spezifisch religionspädagogischer Arbeiten zu KI notwendig, auch kleinere, vorwiegend unterrichtspraktische Publikationen (vgl. 2.3b) heranzuziehen, deren zentrale Aspekte jedoch prägend für die intensivere Forschung der Zukunft werden können. Zum anderen ist darauf hinzuweisen, dass die im Folgenden gewählte Systematik der Darstellung erst *on the long run* auf ihr Potenzial als heuristisches Instrument geprüft werden muss. So wird es kritisch zu reflektieren sein, ob sich das mit hoher Wahrscheinlichkeit stark expandierende und differenzierende Forschungsfeld auch mittelfristig in der vorgeschlagenen Weise ordnen lässt. Unter dem eingepreisten Vorbehalt dieser Vorzeichen wird im Folgenden eine orientierende Kartographie²⁹ religionspädagogischer Forschung zur KI (Erarbeitungsstand 2024) erstellt (Abs. 3.1), um auf dieser Grundlage zentrale Tendenzen religionspädagogischer Forschung zu KI im deutschsprachigen Raum herauszuarbeiten (Abs. 3.2).

²⁷ Vgl. exemplarisch: Ebinger, Johanna / Kaufmann, Sven: Künstliche Intelligenz im Unterricht.

²⁸ Vgl. Chrostowski, Mariusz: Anwendung von Künstlicher Intelligenz im Religionsunterricht, 81; dieser Befund wird sich allerdings wohl bald überholt haben, nachdem KI immer mehr in den Fokus religionspädagogischer Forschung rückt. Vgl. bspw. den Tagungsband: Chrostowski, Mariusz / Heger, Johannes (Hg.): Minding the gap.

²⁹ Begrifflich motiviert ist dieser Überblick durch: Roebben, Bert: Religionspädagogische Kartographie.

Eine orientierende Kartographie religionspädagogischer Forschung zu KI

Da Digitalisierung und KI auf vielfache Weise miteinander verwoben sind, ist es nicht nur möglich, sondern auch geboten, religionspädagogische Forschung zu KI in den bereits bespielten Horizont von Überlegungen zu religiöser Bildung in einer Kultur der Digitalität zu stellen (vgl. Abs. 2).³⁰ Ähnlich zum größeren Horizont religiöser Bildung in einer Kultur der Digitalität lassen sich im Spiegel aktueller religionspädagogischer Forschung mit theologischen und bildungstheoretischen Grundsatzfragen (Abs. 3.1.1), KI in religiösen Bildungsprozessen (Abs. 3.1.2) sowie KI in der Aus- und Fortbildung (Abs. 3.1.3) v. a. drei Schwerpunkte religionspädagogischer Forschung über KI ausmachen.

Theologische und bildungstheoretische Grundsatzfragen

Innerhalb der Religionspädagogik stellen sich – wie auch in ihrem Referenzrahmen der Theologie (vgl. 2.1 b + c) – beim Zugang zu KI zunächst und vor allem anthropologische und ethische Fragen bezüglich des geltenden Menschen- und Weltbildes.³¹ Im Vordergrund stehen dabei insbesondere Aspekte des Trans- und Posthumanismus sowie des Human Enhancement. Einen markanten Hinweis darauf bietet z.B. das Themenheft „Künstliche Intelligenz und Human Enhancement“ der Fachzeitschrift „Religion Unterrichten“ (3. Jahrgang, Heft 2) von 2022,³² in dem u. a. Bernhard Grümme eine religionspädagogische Anthropologie³³ als vorrangiges Anliegen religiöser Bildung im Umgang mit KI herausstellt.³⁴ Weiterhin demonstrieren die in dieses Heft integrierten praxisnahen Beiträge von Heike Harbecke³⁵, Jasmin Hassel³⁶ und Gabriele Otten³⁷, wie man mit Schüler:innen kritisch über KI diskutieren und dabei Verantwortungsbewusstsein auf der Basis einer biblisch-christlichen Überlieferung fördern kann. Auch die „Zeitschrift für Pädagogik und Theologie“ (Band 75, Heft 2) hat im

³⁰ Konsequenterweise wird dies auf institutioneller Ebene bspw. umgesetzt durch den Orientierungsrahmen der EKD für den ev. RU in der digitalen Welt. Vgl. EKD (Hg.): Evangelischer Religionsunterricht in der digitalen Welt, 7.

³¹ Vgl. Puzio, Anna: Theologie und Künstliche Intelligenz: Perspektiven, 20.

³² Vgl. Otten, Gabriele / Paeßens, Jutta: Editorial, 6–7.

³³ Die religionspädagogische Anthropologie nach Bernhard Grümme betrachtet den Menschen als Ganzheit aus Körper, Leib und Seele, als sterbliches, beziehungsorientiertes Geschöpf Gottes. Sie verbindet theologische, philosophische und humanwissenschaftliche Perspektiven, entwickelt universalisierbare Kategorien und reflektiert kritisch anthropologische Annahmen der Religionspädagogik. Geprägt ist sie von einer alteritätstheoretischen Denkweise, die die Bezogenheit des Menschen auf Mitmenschen und Gott betont; vgl. Grümme, Bernhard: Menschen bilden?

³⁴ Vgl. Grümme, Bernhard: KI, ein Ernstfall der Religionspädagogik.

³⁵ Vgl. Harbecke, Heike: Ethisch-anthropologische Herausforderungen selbstfahrender Autos.

³⁶ Vgl. Hassel, Jasmin: Human Enhancement, Big Data und KI.

³⁷ Vgl. Otten, Gabriele: „Ich bin dein Mensch“.

Jahr 2023 ein Sonderheft „Künstliche Intelligenz und Transhumanismus“ auf den Markt gebracht.³⁸ Dieser thematische Trend wird in einem breiteren Kontext u. a. durch weitere wissenschaftliche Arbeiten untermauert: So analysiert bspw. Elisabeth Fock³⁹ die Auswirkungen trans- und posthumanistischer Ideen auf die Wahrnehmung von Körper und Sterblichkeit im Kontext religiöser Bildung; Steffi Fabricius⁴⁰ und Claudia Gärtner⁴¹ entwickeln ferner die Pointe, dass eine theologische oder religionsdidaktische Auseinandersetzung mit Trans- und Posthumanismus nach einer neuen Definition des Menschseins verlangt.

Diese hermeneutischen Tiefenbohrungen werden auch hinsichtlich weiterer thematischer Akzente unternommen: So setzt sich Katharina Zweig aus ethischer Perspektive mit Fragen der Moral und Ethik von KI-Entscheidungen auseinander. Sie kritisiert die Perpetuierung von Stereotypen durch Algorithmen und hinterfragt die moralischen Aspekte der Delegation von Entscheidungen an Maschinen. Außerdem stellt sie fest, dass ein solches Vorgehen dem christlichen Verständnis der Würde des Menschen widerspricht, welches die Grundlage der Subjektorientierung religiöser Bildung bildet.⁴² Ähnlich warnt Birte Platow vor der Gefahr der Identitätsfragmentierung, die sich aus der fortschreitenden Verbreitung digitaler Interaktionen in unserem Alltag ergibt, die das menschliche Selbstbild sowohl im Positiven als auch im Negativen nachhaltig transformieren.⁴³ In diesem erkenntnistheoretischen Zusammenhang verweist Viera Pirker aus christlicher Perspektive auf die besondere Dimension der menschlichen Existenz, die sich einer vollständigen Durchdringung durch technologische Systeme und algorithmische Logik entzieht und als Gegenpol zu den Authentizitätsansprüchen der digitalen Welt verstanden werden kann.⁴⁴

Während solche theologisch konnotierten Fragestellungen nach anthropologisch und ethisch vertretbaren Menschen- und Weltbildern in Bezug auf KI viel Aufmerksamkeit erfahren, finden sich bislang nur wenige Studien, die sich aus religionspädagogischer Perspektive explizit mit bildungstheoretischen Grundsatzfragen im KI-Kontext auseinandersetzen (vgl. Abs. 2.2). Als Ausnahme kann hier u. a. der Text von Kerstin Heinemann genannt werden, in dem Überlegungen zu einer erweiterten Medienbildung vor dem Hintergrund des Aufstiegs von KI-Systemen vertieft werden und für eine transformative (religiöse) Bildung plädiert wird, die den ganzen Menschen als Subjekt lebenslangen Lernens einbeziehen muss.⁴⁵ Insofern stellt KI in der religionspädagogischen Literatur bildungstheoretisch eher eine Marginalie

³⁸ Vgl. Koerrenz, Ralf / Schlag, Thomas: Editorial, 125–127; weitere Hefte in ähnlicher Absicht sind in Planung bzw. im Erscheinen (vgl. bspw. Entwurf 1/2025; Katechetische Blätter 1/2026).

³⁹ Vgl. Fock, Elisabeth: Der Körper im Religionsunterricht.

⁴⁰ Vgl. Fabricius, Steffi: Transhumanismus.

⁴¹ Vgl. Gärtner, Claudia: Digitales ‚Ich‘?

⁴² Vgl. Zweig, Katharina: Künstliche Intelligenz im Religionsunterricht.

⁴³ Vgl. Platow, Birte: Selbstwahrnehmung und Ich-Konstruktion im Angesicht von Künstlicher Intelligenz, 105–124; Platow, Birte: Digitalisierung / Big Data / Künstliche Intelligenz.

⁴⁴ Vgl. Pirker, Viera: Das Geheimnis im Digitalen. Anthropologie und Ekklesiologie im Zeitalter von Big Data und Künstlicher Intelligenz, 133–141; Pirker, Viera, Subjekt Mensch.

⁴⁵ Vgl. Heinemann, Kerstin: „Die sind nicht ganz so schnell, die ganzen Algorithmen.“

dar, die gewissermaßen als komplementäre Dimension auftritt, wenn das Problem der religiösen Bildung in der digitalen Welt thematisiert wird.⁴⁶

KI als Lerngegenstand und Tool in religiösen Bildungsprozessen

Mit Blick auf KI als Lerngegenstand und Tool in religiösen Bildungsprozessen lässt sich zunächst feststellen, dass sich die meisten Publikationen mit religionsdidaktischen Fragen beschäftigen (vgl. 2.3). Ein Annäherungsversuch zu diesem Thema stammt u. a. von Mariusz Chrostowski. In seinem 2023 veröffentlichten Aufsatz „Anwendung von Künstlicher Intelligenz im Religionsunterricht“⁴⁷ untersucht er den Einsatz von KI zur Unterstützung der Kernaufgaben des schulischen Religionsunterrichts, „d. h. im Rahmen der Vermittlung von strukturiertem und lebensbedeutsamem Grundwissen über den Glauben der Kirche; des Vertrautmachens mit Formen gelebten Glaubens und der Förderung religiöser Dialog- und Urteilsfähigkeit.“⁴⁸ Er beleuchtet sowohl die Chancen als auch die Grenzen der Implementierung von KI in religionsunterrichtlichen Settings, indem er u. a. darauf hinweist, dass KI einerseits neue Zugänge zu Religion und den damit verbundenen Traditionsbeständen eröffnen, andererseits aber auch kontraintentional die Subjektwerdung der Schüler:innen untergraben und Fehlinformationen und Vorurteile reproduzieren kann.⁴⁹ Ein weiterer Text aus seiner Feder greift die Frage der KI als Lerngegenstand im Religionsunterricht sowie die mit den neuen Technologien einhergehenden anthropologischen, ethischen und theologischen Herausforderungen auf.⁵⁰ Zudem diskutiert Chrostowski in einem anderen Beitrag die Förderung kritischen Denkens durch den Einsatz von ChatGPT im Religionsunterricht und argumentiert, dass kritisches Denken, welches per se zur religiösen Bildung gehört, dem unterrichtlichen und außerunterrichtlichen Einsatz von KI-Tools immer vorausgehen muss.⁵¹ In Zusammenarbeit mit Andrzej J. Najda konzentriert sich Chrostowski in einem nächsten Fachartikel auf die Nutzung von Chatbots im biblischen Lernen und lotet dabei die Möglichkeiten von ChatGPT anhand des Fünf-Ebenen-Modells der Bibeldidaktik von J. Theis aus.⁵²

⁴⁶ Vgl. u. a.: Nord, Ilona / Pirner, Manfred: Religiöse Bildung in der digitalen Welt; Gronover, Matthias u. a. (Hg.): Religiöse Bildung in einer digitalisierten Welt; Tomberg, Markus / Verburg, Winfried (Hg.): RU 4.0 Religiöse Bildung und Digitalisierung; Simojoki Henrik: Digitalisierung – Herausforderung und Chance für den Religionsunterricht.

⁴⁷ Vgl. Chrostowski, Mariusz: Anwendung von Künstlicher Intelligenz im Religionsunterricht.

⁴⁸ Ebd., 79.

⁴⁹ Vgl. ebd., 87–89.

⁵⁰ Chrostowski, Mariusz: Anthropologische, ethische und theologische Problemzonen der Künstlichen Intelligenz als Lerngegenstand im Religionsunterricht.

⁵¹ Vgl. Chrostowski, Mariusz: Förderung kritischen Denkens durch den Einsatz von ChatGPT im Religionsunterricht.

⁵² Vgl. Chrostowski, Mariusz / Najda, Andrzej: ChatGPT as a modern tool for Bible teaching in confessional religious education.

Johannes Heger bereichert das Feld mit mehreren Akzentsetzungen: In seinen Publikationen über KI wirbt er u. a. für die Relevanz, in Forschung, Lehre und Bildungsprozessen unreflektierte Haltungen gegenüber der Technologie selbstkritisch zu dekonstruieren.⁵³ Er macht sich stark für ein Verständnis religiöser Bildung, die das Subjekt bei seiner Selbstwerdung in der Kultur der Digitalität unterstützt, indem KI in religiöse Lern- und Bildungsprozesse integriert wird⁵⁴ und sieht KI aus bildungstheoretischen und theologischen Gründen als notwendigen Bezugspunkt religiöser Bildung.⁵⁵ Dies umfasst für Heger sowohl einen Umgang mit KI als Tool⁵⁶ als auch eine theologische bzw. ethische Auseinandersetzung mit KI als Lerngegenstand. Eine besondere Aufmerksamkeit widmet er bei seinen Überlegungen dem Potenzial, KI im Kontext der Eschatologie(didaktik) zu bedenken.⁵⁷

Ebenfalls mit KI als Unterrichtsgegenstand und Werkzeug (z. B. zur Erstellung von Bildern zu biblischen Gleichnissen oder zur Unterstützung bei der Informationsbeschaffung zu religiösen Themen) befassen sich u. a. die Studien von Kathrin Termin und Simon Kluge.⁵⁸ Erwähnenswert ist in diesem Zusammenhang auch die kürzlich erschienene Masterarbeit von Sandra Kunz, die die Möglichkeiten der Anwendung von KI im Kontext der beruflichen Bildung untersucht. Dabei liegt ihr Fokus auf dem spezifischen Potenzial des Religionsunterrichts, einen mündigen Umgang der Lernenden mit neuen Technologien zu fördern. Neben ethischen und theologischen Aspekten der Digitalisierung zeigt Kunz in ihrer Arbeit konkrete Szenarien auf, wie KI im schulischen Alltag sinnvoll zum Tragen kommen kann.⁵⁹

Hervorzuheben sind ferner religionspädagogische Arbeiten, die sich mit den Themen Künstliche Intelligenz, Metaversum und Virtuelle Realität (VR) auseinandersetzen. Dazu gehört insbesondere der Aufsatz von Thomas Schlag und Katharina Yadav, in dem das Metaversum als potenziell bedeutsamer, KI-basierter und immersiver Raum für den Religionsunterricht skizziert wird. Die Autoren betonen dabei aus christlicher Perspektive die Rolle des Metaversums für die Persönlichkeitsentwicklung, die individuelle Orientierung sowie für Erfahrungen von Körperlichkeit, Beziehungs- und Resonanz Erfahrungen und verantwortliches Handeln im Religionsunterricht.⁶⁰

Weitere religionspädagogische Publikationen beschäftigen sich mit KI-basierten sozialen Robotern in religiösen Lern- und Lehrprozessen. Steffi Fabricius betont die Rolle von

⁵³ Vgl. Heger, Johannes: Künstliche Intelligenz und religiöse Bildung, v. a. 39f.

⁵⁴ Vgl. Heger, Johannes / Leven, Eva: Religiöse Bildung in einer Kultur der Digitalität, v. a. 18f.; Ders.: Digital, medial, egal?, v. a. 43–49.

⁵⁵ Vgl. Heger, Johannes: (Religiöse) Bildung mit KI in der Zukunft.

⁵⁶ Vgl. Heger, Johannes: Heilige Lade oder Golem?

⁵⁷ Vgl. Heger, Johannes: Künstliche Intelligenz – ein Endzeitbeschleuniger?; Ders.: Eschatologie, v. a. Abs. 5.1; Heger, Johannes: Von Visionen, Halluzinationen und Rationalität.

⁵⁸ Vgl. Termin, Kathrin: KI – besser als der Mensch?; Dies.: KI-Tools mit Schüler: innen nutzen; Kluge, Simon: Künstliche Intelligenz als Thema im Religionsunterricht.

⁵⁹ Kunz, Sandra: KI-gestützte Lernwelten.

⁶⁰ Vgl. Schlag, Thomas / Yadav, Katharina: Bildungsbeziehungen in der Ankunftszeit des Metaversums.

Robotern als „digitale Glaubenszeugen“⁶¹ und eröffnet damit neue Perspektiven für deren Einsatz in der Religionsdidaktik, während Mariusz Chrostowski in seiner theoretischen Annäherung eine Konzeptualisierung interaktiver Lehr-Lernformen mit sozialen Robotern im Religionsunterricht der Zukunft vornimmt.⁶²

Des Weiteren ist zu konstatieren, dass Religionspädagog:innen fortlaufend Unterrichtsentwürfe und Praxishilfen konzipieren, die sich thematisch und methodisch mit KI auseinandersetzen. Beispielhaft seien hier die Praxismaterialien von Jens Palkowitsch-Kühl und Eva-Maria Leven⁶³, Jochen Walldorf⁶⁴, Simone Meinen⁶⁵ oder der Materialpool von rpi-virtuell⁶⁶ und ReliBoard⁶⁷ genannt (vgl. Abs. 2.3b).

KI als Lerngegenstand und Tool in der Aus- und Fortbildung von Religionslehrkräften

Umfassende Veröffentlichungen, die sich gezielt mit KI als Lerngegenstand und Werkzeug in der Aus- und Fortbildung von Religionslehrkräften beschäftigen, sind bisher kaum vorhanden. Die wenigen existierenden Arbeiten sind meist fragmentarische Analysen mit einem idographischen Ansatz. In gewissem Maße greift der bereits erwähnte Beitrag von Johannes Heger das Thema auf: Er skizziert hochschul- und religionsdidaktische Überlegungen zum Einsatz von ChatGPT und beleuchtet dabei u. a. Möglichkeiten für einen kriteriengeleiteten Umgang mit Chatbots in der universitären Lehre und in Prüfungsformaten.⁶⁸ Einen weiteren Blickwinkel bietet die Publikation von Ilona Nord und Jens Palkowitsch-Kühl, in der die Autoren darauf hinweisen, dass die digitale Kommunikation die didaktischen Arrangements modifiziert und damit eine Professionalisierung der Religionslehrer:innen erfordert.⁶⁹ Theresia Witt et al. befassen sich mit innovativen Formaten in digitalen Bildungskulturen der Hochschullehre,⁷⁰ während Konstantin Lindner betont, dass Religionslehrkräfte spezifische Kompetenzen im Bereich der digitalen Souveränität benötigen. Dazu gehört nicht nur die Fähigkeit, digitale Medien selbstbestimmt zu nutzen, sondern auch die Entwicklung einer flexiblen Haltung gegenüber den Herausforderungen einer zunehmend digitalisierten Welt.

⁶¹ Vgl. Fabricius, Steffi: Robocatio.

⁶² Vgl. Chrostowski, Mariusz: Zum Einsatz sozialer Roboter im Religionsunterricht.

⁶³ Vgl. Palkowitsch-Kühl, Jens / Leven, Eva-Maria: Wie begegnen wir künstlicher Intelligenz?.

⁶⁴ Vgl. Walldorf, Jochen: Mensch und Maschine.

⁶⁵ Vgl. Meinen, Simone: KI im RU.

⁶⁶ Vgl. Rpi-virtuell: Künstliche Intelligenz.

⁶⁷ Vgl. ReliBoard: Unterrichtsideen und Tipps.

⁶⁸ Vgl. Heger, Johannes: Heilige Lade oder Golem?

⁶⁹ Vgl. Nord, Ilona / Palkowitsch-Kühl, Jens: Nicht die App steht im Mittelpunkt, sondern der Kompetenzerwerb.

⁷⁰ Vgl. Witt, Theresia u. a. (Hg.): Diversität und Digitalität in der Hochschullehre.

Dabei wird jedoch noch nicht explizit auf die spezifischen Anforderungen im Bereich der Künstlichen Intelligenz eingegangen.⁷¹ Ähnlich verhält es sich bei Katharina Grundens empirischer Studie, mit der sie die Wirksamkeit digitalen Lehrens und Lernens im Vergleich zu analogen Lehr-Lernformaten untersucht hat.⁷²

Darüber hinaus verdient das laufende Dissertationsprojekt von Kathrin Termin Erwähnung, auch wenn es sich nicht explizit mit KI beschäftigt. Unter dem Titel „Professionalisierung von Religionslehrkräften auf dem Feld der Digitalisierung – eine empirische Begleitforschung“ untersucht sie die Motive von Religionslehrkräften, sich im Bereich der digitalen Bildung zu professionalisieren, und die Rolle des ‚relilab‘ als digitalen Raum für ihre Weiterbildung.⁷³

Tendenzen und Forschungsdesiderate im Themenfeld KI

Dieser kursorisch skizzierte state of the art religionspädagogischer Forschung zum Themenfeld KI zeigt, dass im Bereich der evangelischen und katholischen Religionspädagogik derzeit zahlreiche fachspezifische Beiträge zum Thema KI entstehen. Die analysierten Pionierarbeiten konzentrieren sich häufig auf spezifische Teilaspekte des breiten Themenfeldes KI. Dabei artikulieren grundlagentheoretisch akzentuierte Ansätze, wie etwa jene von Bernhard Grümme, eher Bedenken und markieren die theologisch-ethischen Grenzen von KI im Hinblick auf die christliche Anthropologie.⁷⁴ Andere Positionen, wie sie z. B. von Mariusz Chrostowski, Johannes Heger, Simon Kluge und Kathrin Termin vertreten werden, versuchen dagegen, KI vor allem als Hilfsmittel oder Lerngegenstand in der religionsdidaktischen Praxis zu implementieren. Sie betonen eine Reihe von pragmatischen Vorteilen wie die Förderung eines reflektierten Umgangs der Lernenden mit digitalen Medien, die stärkere Integration der KI und die dadurch mögliche Optimierung von Bildungsprozessen, die Entlastung der Lehrkräfte etc.⁷⁵

Während die theologisch und religionsdidaktisch orientierten Zugänge zu KI erste relevante Grundlagen für religiöse Bildungsprozesse schaffen, bleibt die Frage weitgehend offen, wie religiöse Bildung insgesamt angesichts der Herausforderungen und Potenziale von KI neu gedacht und konzeptionell weiterentwickelt werden kann. Es fehlen umfassende und integrative Modelle, wie dies z. B. von Sandra Kunz partiell versucht wurde,⁷⁶ welche die Möglichkeiten und Grenzen von KI in der Breite religiöser Bildungsprozesse systematisch reflektieren

⁷¹ Vgl. Lindner, Konstantin: Professionalität von Religionslehrer:innen.

⁷² Vgl. Grundens, Katharina: Das Ende der „Kreidezeit“.

⁷³ Vgl. Relilab: Begleitende Forschung.

⁷⁴ Vgl. Grümme, Bernhard: KI, ein Ernstfall der Religionspädagogik.

⁷⁵ Vgl. Chrostowski, Mariusz: Anwendung von Künstlicher Intelligenz im Religionsunterricht; Heger, Johannes: Künstliche Intelligenz und religiöse Bildung; Kluge, Simon: Künstliche Intelligenz; Termin, Kathrin: KI-Tools mit Schüler:innen nutzen.

⁷⁶ Vgl. Kunz, Sandra: KI-gestützte Lernwelten.

und praktisch nutzbar machen. Genau hier zeigt sich ein dringendes Forschungsdesiderat, nämlich die Ausarbeitung bildungstheoretisch fundierter und empirisch überprüfter Konzepte religiöser Bildung angesichts von KI.

In diesem Sinne müssten im Bereich der Religionspädagogik zentrale, bislang unzureichend erforschte Felder im Hinblick auf KI gezielt adressiert werden. Dazu gehören etwa die Rolle neuer Technologien bei der Selbstoptimierung, aber auch bei der Unterstützung von Prozessen der Subjektwerdung und (christlichen) Identitätsbildung, die Förderung nachhaltiger Religiosität, die Interpretation und Kontextualisierung religiöser Narrative, die Analyse und kritische Auseinandersetzung mit Gottesbildern und Weltdeutungen sowie die Förderung (inter-)religiöser Kompetenz etc. im Horizont der digitalen Transformation. Damit einhergehen sollte die Frage nach der Professionalisierung von Religionslehrer:innen, die – wie u. a. Thomas Schlag und Katharina Yadav zeigen – im Kontext von KI, Metaversum und Virtual Reality nicht nur technische Fähigkeiten, sondern auch kritische Reflexionskompetenz und sinnvolle Bildungsbeziehungen entwickeln müssen.⁷⁷ Gerade im Bereich der Lehrerbildung könnte eine umfassendere Auseinandersetzung mit digitaler Souveränität entscheidende Impulse geben. Religionslehrkräfte stehen – wie z. B. Konstantin Lindner sowie I. Nord und J. Palkowitsch-Kühl zutreffend betonen – vor der Herausforderung, ihre Rolle in einer zunehmend digitalisierten Bildungslandschaft zu überdenken und sich aktiv sowie kompetent mit den Potenzialen und Risiken neuer Technologien auseinanderzusetzen. Dies erfordert nicht nur technisches Wissen, sondern auch ein vertieftes Verständnis kultureller, sozialer und religiöser Implikationen von KI, was sich auch in der Ausbildung von Lehrkräften niederschlagen sollte.⁷⁸

Vor dem Hintergrund aktueller Entwicklungen in der Religionspädagogik lassen sich einige exemplarische Forschungsdesiderate formulieren, die sowohl hermeneutische als auch empirische Aspekte berücksichtigen sollten, insbesondere in Bezug auf die Frage, inwieweit KI in religiöse Lern- und Lehrsettings integriert werden kann. Als konzeptionelle Grundlage bieten sich z. B. die zentralen fachspezifischen Bereiche religiöser Bildung aus dem „Handbuch Religionsdidaktik“ von Ulrich Kropač und Ulrich Riegel an: die Akteure im Religionsunterricht sowie die didaktischen Lernformen.⁷⁹ Darauf aufbauend können unter anderem folgende Fragestellungen in künftigen religionsdidaktischen bzw. religionspädagogischen Arbeiten in den Blick genommen werden:

- a) bezogen auf die Personen des Religionsunterrichts: Die religiös-weltanschauliche Entwicklung der Lernenden könnte dahingehend untersucht werden, wie KI deren individuelle religiöse Entwicklung fördert oder vereinfacht. Im Hinblick auf die religiöse Identitätsbildung der Schüler wäre es relevant, zu erforschen, inwieweit KI personalisierte Lernwege unterstützt, ohne die vielschichtige Natur der Identitätsbildung zu

⁷⁷ Vgl. Schlag, Thomas / Yadav, Katharina: Bildungsbeziehungen in der Ankunftszeit des Metaversums.

⁷⁸ Vgl. Lindner, Konstantin: Professionalität von Religionslehrer:innen; Nord, Ilona / Palkowitsch-Kühl, Jens: Nicht die App steht im Mittelpunkt, sondern der Kompetenzerwerb.

⁷⁹ Vgl. Kropač, Ulrich / Riegel, Ulrich (Hg.): Handbuch Religionsdidaktik.

unterlaufen. Ebenso sollten die Auswirkungen von KI auf die berufliche Entwicklung und Unterrichtsqualität von Lehrpersonen untersucht werden, insbesondere im Hinblick auf die Akzeptanz und Integration neuer Technologien.

- b) bezogen auf religionsdidaktische Lernformen: Bezüglich des Lernens mit Symbol-Zeichen wäre zu untersuchen, wie KI das Verständnis komplexer religiöser Symbole und Zeichen beeinflusst, wobei die Herausforderung in der diesbezüglichen Präzision und religiösen Sensibilität der KI-Modelle liegt. Beim ökumenischen und interreligiösen Lernen könnte erforscht werden, wie KI den interkonfessionellen Austausch unterstützt, ohne Vorurteile zu verstärken und Missverständnisse zu erzeugen. In Bezug auf ethisches Lernen wäre zu prüfen, wie KI ethische Dilemmata realistisch simulieren und reflektieren kann, wobei die Komplexität moralischer Fragestellungen berücksichtigt werden muss. Im Bereich des biblischen Lernens könnte untersucht werden, wie KI interaktives Lernen unterstützt, ohne fehlerhafte Interpretationen und Kontextualisierungen zu riskieren. Weiterhin könnte erforscht werden, inwieweit KI Perspektivübernahme und -wechsel in der religiösen Bildung fördert, wobei die technische und didaktische Umsetzung komplexer religiöser Lernprozesse eine Herausforderung darstellt. Schließlich sollte ausgelotet werden, auf welche Weise KI das Lernen und Lehren an außerschulischen religiösen Lernorten verbessern kann, ohne die Qualität religiöser Bildung und zwischenmenschlicher Interaktionen zu beeinträchtigen.

Neben diesen differenzierten, spezifisch religionspädagogischen Überlegungen darf – die bereits gewordene Praxis (vgl. 2.3b) im Blick – nicht unterschlagen werden, dass die religiöse Bildungspraxis auch hinsichtlich fächerübergreifender Aspekte unterrichtlichen Handelns bereichert wird: So profitieren auch Religionslehrkräfte von KI – bspw. bei der Unterrichtsvorbereitung, bei der Bewältigung organisatorischer Aufgaben und durch die Bereitstellung differenzierter Materialien.⁸⁰ Bei aller Effizienzsteigerung stößt sie jedoch an ihre Grenzen, wenn es um spezifische religionspädagogische Fragestellungen geht. Subjektorientiertes Lernen und Sinn- bzw. Werteorientierung in religiösen Bildungsprozessen setzen die Expertise und pädagogische Sensibilität der Religionslehrer:innen voraus, da diese Aspekte von der KI allein nicht geleistet werden können.⁸¹ Zugleich ist zu berücksichtigen, dass viele der

⁸⁰ „Ein Blick in die methodische Werkzeugkiste von Religionslehrpersonen und Katechet*innen – z. B. die Online-Sammlungen von Arthur Thömmes („Künstliche Intelligenz im Unterricht. Tools, Informationen und Beispiele“) oder Simone Meinen („KI im Religionsunterricht“) – lässt nicht nur erkennen, dass KI bereits Einzug in religiöse Bildungspraxis gehalten hat, sondern dass KI hier in Form von bestimmten, mehr oder weniger einfach zugänglichen Tools Einsatz findet. Hierzu zählen u. a. Textgeneratoren (z. B. ChatGPT, Chatbots wie Nikodemus.AI vom ERF Bibleserver), Bildgeneratoren (z. B. DallE), Musikgeneratoren (z. B. Suno) sowie Präsentationsgeneratoren (z. B. mysimpleshow). Die Palette an Werkzeugen wächst mit jedem Tag, manche Tools kommen hinzu, andere verschwinden. Den Überblick zu behalten, ist selbst für Geübte oft schwierig“; Mayrhofer, Florian: Von Jesus-Bots und Luther-Avataren, 2.

⁸¹ Vgl. Chrostowski, Mariusz: Anwendung von Künstlicher Intelligenz im Religionsunterricht, 87–89.

derzeit kursierenden allgemeinen Handreichungen im Hinblick auf KI adaptierbar sind,⁸² während sich die spezifischen religionsdidaktischen Anforderungen erst in der praktischen Umsetzung herauskristallisieren.

Ferner ist festzustellen, dass die religionspädagogische Forschungslandschaft von einer stärkeren interdisziplinären Perspektive profitieren könnte, wie sie bereits in der Medienpädagogik und den Bildungswissenschaften etabliert ist. So könnten z. B. zentrale Aspekte wie kritische KI-Literalität und digitale Ethik stärker in religiöse Bildungsprozesse integriert werden.⁸³ Dies lässt sich u. a. auf zwei Ebenen umsetzen: Einerseits sollte an Schulen und Hochschulen mehr Freiraum für experimentelle Erprobung und kritische Reflexion im Hinblick auf KI geschaffen werden, sodass die daraus gewonnenen Erkenntnisse iterativ in die religionspädagogische Praxis zurückfließen können. Andererseits ist die Religionspädagogik in enger Kooperation mit anderen Fächern explizit gefordert, neue Formate zur Effektivierung des Theorie-Praxis-Verhältnisses zu entwickeln, insbesondere hinsichtlich der Lern- und Lehrwirksamkeit sowie der Gestaltung neuer Bildungsszenarien im Religionsunterricht und deren kritischer Begleitung. Diese Aufgaben müssen theoriebasiert und präziser als bisher angegangen werden.⁸⁴ Zudem ist es angesichts der teils neuen, teils bekannten, aber durch die rasante (R)Evolution der KI verstärkten Herausforderungen notwendig, interdisziplinär zu hinterfragen, welche Ziele das Bildungssystem aktuell verfolgt und an welchen Zielen es sich in Zukunft überhaupt orientieren sollte.⁸⁵ In diesem Zusammenhang liegt der Schwerpunkt der bisherigen Arbeiten auf hermeneutisch-konzeptionellen Ansätzen und Überlegungen. Eine empirische Untersuchung der Zusammenhänge, wie sie teilweise bereits für den Klangraum der digitalen Kultur vorliegt, wäre jedoch wünschenswert (vgl. Abs. 3.1.2 und 3.1.3).

Gretchen, KI und Religionspädagogik II

Eingangs wurde im Rückgriff auf Goethes Faust mit der Frage operiert, ob die Religionspädagogik nicht einen Tanz mit dem Teufel wagt, wenn sie sich des Themenfeldes KI annimmt (Abs. 1). Die zurückliegenden Ausführungen geben auf diese pointierte Frage zwar keine unmittelbare Antwort. Allerdings eröffnen sie die Möglichkeit, mittelbar eine direkte Antwort aufzubauen.

⁸² Vgl. u. a.: Evangelische Kirche im Rheinland, Handreichung; Caruso, Carina / Hengesbach, Rudolf: Angehende (Religions-)Lehrkräfte begleiten.

⁸³ Vgl. u. a.: Kuka, Lisa u. a.: Menschliche Gestaltung; Köstler, Verena: Zwischen Präzision und Sensitivität; Klar, Maria / Schleiss, Johannes: Künstliche Intelligenz im Kontext von Kompetenzen, Prüfungen und Lehr-Lern-Methoden; Knaus, Thomas: Warum KI kein Hype ist und die Medienpädagogik sich damit befassen sollte.

⁸⁴ Vgl. Klar, Maria / Schleiss, Johannes: Künstliche Intelligenz im Kontext von Kompetenzen, Prüfungen und Lehr-Lern-Methoden, 52.

⁸⁵ Vgl. ebd.

Entlang der vorliegenden Abhandlung könnte diese folgendermaßen formuliert werden: Wenn die Religionspädagogik ihrem eigenen wissenschaftstheoretischen Selbstbild entsprechen will, dann muss sie sich als dynamische, auf Lebenswelt bezogene Disziplin der neuen Technologie und ihrer Implikationen für religiöse Bildung annehmen (Abs. 2). Durch ihre disziplinären Referenzrahmen der Theologie (Abs. 2.1) und der Bildungswissenschaften (Abs. 2.2) sowie durch ihren Bezug zum religionsdidaktischen Referenzrahmen (Abs. 2.3) ist ihre Erschließung von KI spezifisch bzw. originär: Sie beforscht KI im Kontext religiöser Bildung und kann damit einen domänenspezifischen Beitrag zur KI- bzw. Medienkompetenz bzw. zum interdisziplinären Diskurs über KI leisten. Dieses theoretische Potenzial wird faktisch noch nicht voll ausgeschöpft (Abs. 3.2). Allerdings zeigen die vorliegenden religionspädagogischen Publikationen bereits deutlich ein Profil: Im Spiegel des Erarbeiteten wird klar, wie sehr es lohnt, die Erschließungskraft der Theologie (Abs. 2.1) intensiver für (religiöse) Bildung auszuloten (Abs. 3.1.1) und weitere Möglichkeiten zu entdecken, wie KI sowohl als Lerngegenstand und Tool auf der Ebene des Religionsunterrichts (Abs. 3.1.2) als auch in der Professionalisierung von Religionslehrkräften (Abs. 3.1.3) eingesetzt werden kann.

Diese deskriptive Conclusio lässt sich abschließend auf die KI-Gretchenfrage hin zuspitzen – und das mit zwei pointierten Gedankenspielen und auf diesen ruhenden Thesen:

(a) Gelänge es der Religionspädagogik, die nunmehr kondensierten Potenziale religiöser Bildung angesichts Künstlicher Intelligenz überzeugend zu heben, dann wage sie keinen Tanz mit dem Teufel. Vielmehr erweise die Religionspädagogik als Disziplin und religiöse Bildung als Element der Allgemeinbildung, dass sie wahrhaft an der Lebenswelt und den Lernenden Subjekten orientiert ist. Ließe sie sich dagegen nicht auf diese Auseinandersetzung ein, dann bliebe die Frage, ob Schule und Gesellschaft in einer Kultur der Digitalität überhaupt noch langfristig den Tanz mit Religion, Theologie und Religionsunterricht wagen wollen.

Daraus folgt These I: Wenn religiöse Bildung gesellschaftlich und (bildungs-)politisch relevant bleiben will, dann muss sie den Tanz mit der (teuflischen?) KI wagen!

(b) Neigte man dazu, in KI tatsächlich den Teufel zu sehen, dann wäre diese Hinwendung zur religionspädagogischen Erschließung von KI sogar noch alternativloser. Denn nur Lernende, die um die vermeintlich diabolischen Aspekte der Technologie wissen, könnten diese reflektiert bearbeiten.

Daraus folgt These II: Ungeachtet dessen, ob eher die Gefahren oder Chancen der KI Antriebsfeder des (bildungs-)politischen bzw. (religions-)didaktischen Handelns sind; beide Motivationen führen dazu, (religiöse) Bildung mit und über KI intensiver zu beforschen und ihre praktische Implementierung in Curricula voranzutreiben.

Mit diesen Thesen ist der eröffnete Spannungs- und Fragebogen abgerundet und die selbst gestellten Aufgaben sind abgegolten. Die Schnelligkeit der technologischen Entwicklungen bedingt allerdings per se, dass die nunmehr formulierten Erkenntnisse ebenso rasch einer Aktualisierung unterzogen werden müssen. Aber auch dieser finale Gedanke verweist auf einen Ausgangspunkt: Die Dynamik ist ein Wesenskern der Religionspädagogik (vgl. Abs. 2.) – nicht erst seit der Erfolgsgeschichte Künstlicher Intelligenz!

Literatur

- Blossfeld, Hans-Peter u. a.: Digitale Souveränität und Bildung. Gutachten hg. von der vbw, Münster 2018.
- BMBF (Hg.): BMBF-Aktionsplan Künstliche Intelligenz. Neue Herausforderungen chancenorientiert angehen, Berlin 2023.
- Boschki, Reinhold: Der phänomenologische Blick: „Vierschritt“ statt „Dreischritt“ in der Religionspädagogik, in: Gronover, Matthias/Boschki, Reinhold (Hg.): Der phänomenologische Blick: „Vierschritt“ statt „Dreischritt“ in der Religionspädagogik, (= Tübinger Perspektiven zur Pastoraltheologie und Religionspädagogik, Bd. 31), Berlin 2007, 25–47.
- Caruso, Carina / Hengesbach, Rudolf, Angehende (Religions-)Lehrkräfte begleiten, aber wie? Eine Handreichung für die Ausbildung von Lehrkräften sowie für die Praxis in Schule und Unterricht, Paderborn 2021.
- Chrostowski, Mariusz / Najda, Andrzej. J.: ChatGPT as a modern tool for Bible teaching in confessional religious education: a German view, *Journal of Religious Education* (2024), 1–20.
- Chrostowski, Mariusz/Heger, Johannes (Hg.): „Minding the Gap“. Religiöse Bildung angesichts Künstlicher Intelligenz, Freiburg 2026 (i. E.).
- Chrostowski, Mariusz: Anwendung von Künstlicher Intelligenz im Religionsunterricht: Möglichkeiten, Grenzen und Brennpunkte, in: *Religionspädagogische Beiträge* 46 (2023), 1, 79–95.
- Chrostowski, Mariusz: Auf dem Weg zur ‚AI literacy‘ im Religionsunterricht: Versuch einer religionsdidaktischen Konzeptualisierung, in: *Zeitschrift für Pädagogik und Theologie* 77 (2025) 4, 394–412.
- Chrostowski, Mariusz: Förderung kritischen Denkens durch den Einsatz von ChatGPT im Religionsunterricht: Sinn oder Unsinn?, in: *Österreichisches Religionspädagogisches Forum* 32 (2024), 2, 166–181.
- Chrostowski, Mariusz: XIV – 5.1.2.7 Anthropologische, ethische und theologische Problemzonen der Künstlichen Intelligenz als Lerngegenstand im Religionsunterricht, in: Klöcker, Michael / Two-ruschka, Udo / Rötting, Martin (Hg.): *Handbuch der Religionen: Kirchen und andere Glaubensgemeinschaften in Deutschland und im deutschsprachigen Raum*, 82, Hohenwarsleben 2024, 1–20.
- Chrostowski, Mariusz: Zum Einsatz sozialer Roboter im Religionsunterricht: Theoretische Annäherung an interaktive Lehr-Lernformen der Zukunft, in: *Theo-Web: Zeitschrift für Religionspädagogik* 23 (2024), 2, 295-309.
- Deutscher Ethikrat: Mensch und Maschine – Herausforderungen durch KI. Stellungnahme, Berlin 2023.
- Ebinger, Johanna / Kaufmann, Sven: Künstliche Intelligenz im Unterricht. Sprachgesteuerte KI praxisorientiert einsetzen, Berlin 2023.
- EKD (Hg.): Evangelischer Religionsunterricht in der digitalen Welt Ein Orientierungsrahmen (EKD-Texte 140), in: https://www.ekd.de/ekd_de/ds_doc/ekd-texte_140_2022.pdf (abgerufen am 26.10.2024).
- Evangelische Kirche im Rheinland, Handreichung, in: <https://handreichung-ev-ru.de/> (abgerufen am 16.12.2024).

- Fabricius, Steffi: Robocatio: Digitale Zeugen des Glaubens. Roboter und KI im modernen Religionsunterricht, in: Brand, Lukas / Kutz, Martin / Winter, Dominik (Hg.): Instrumente Gottes. Praktische und systematische Reflexionen zum Einsatz sozialer Roboter und KI-Tools in religiösen Kontexten, Baden-Baden 2024 (in print).
- Fabricius, Steffi: Transhumanismus, in: https://cms.ibep-prod.com/app/uploads/sites/18/2023/08/Transhumanismus__2021-02-03_12_26.pdf (abgerufen am 28.10.2024).
- Fock, Elisabeth: Der Körper im Religionsunterricht. Religionspädagogische Vergewisserungen in systematisierender Absicht, in: Journal für allgemeine Didaktik: JfAD 12 (2024), 63–83.
- Gärtner, Claudia: Digitales ‚Ich‘? Trans- und Posthumanismus als theologische und religionspädagogische Herausforderung, in: Konz, Britta / Ostmeyer, Karl-Heinrich / Scholz, Marcel (Hg.): Gratwanderung künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI, Stuttgart 2022, 111–124.
- Glasze, Georg / Odzuck, Eva / Staples, Ronald (Hg.): Was heißt digitale Souveränität? Diskurse, Praktiken und Voraussetzungen „individueller“ und „staatlicher Souveränität“ im digitalen Zeitalter (= Politik in der digitalen Gesellschaft 3), Bielefeld 2022.
- Gronover, Matthias / Obermann, Andreas / Schnabel-Henke, Hanne (Hg.): Religiöse Bildung in einer digitalisierten Welt. Beiträge zur Theorie und Praxis des Religionsunterrichts an berufsbildenden Schulen (Glaube – Wertebildung – Interreligiosität 23), Münster 2021.
- Grümme, Bernhard, Menschen bilden? Eine religionspädagogische Anthropologie, Freiburg 2012.
- Grümme, Bernhard: KI, ein Ernstfall der Religionspädagogik, in: Religion unterrichten 2 (2022), 38–43.
- Grunden, Katharina: Das Ende der „Kreidezeit“ – geht gute Lehre nur digital? Eine qualitative Studie über den Zusammenhang von digitalen Medien und Lernerfolg im Fachbereich Religionspädagogik (= Empirische Theologie 35), Berlin 2022.
- Haidt, Jonathan: Generation Angst. Wie wir unsere Kinder an die virtuelle Welt verlieren und ihre psychische Gesundheit aufs Spiel setzen, Hamburg 2024.
- Harari, Yuval Noah: Nexus. Eine kurze Geschichte der Informationsnetzwerke von der Steinzeit bis zur künstlichen Intelligenz, München 2024.
- Harbecke, Heike: Ethisch-anthropologische Herausforderungen selbstfahrender Autos – mit Gestaltungsaufgabe(n) Bausteine für ein Unterrichtsvorhaben ab der Jahrgangsstufe 10, in: Religion unterrichten 2 (2022), 44–70.
- Hasenbein, Melanie: Der Mensch im Fokus der digitalen Arbeitswelt. Wirtschaftspsychologische Perspektiven und Anwendungsfelder, Berlin 2020.
- Hassel, Jasmin: Human Enhancement, Big Data und KI – ethische Herausforderungen im RU. Bausteine für die Jahrgangsstufen 10–12, in: Religion unterrichten 2 (2022), 71–100.
- Heger, Johannes/Leven, Eva: Religiöse Bildung in einer Kultur der Digitalität – Grundsätzliche Ausrichtungen und Optionen religiösen Lernens mit KI, in: IRP – Impulse 2/2024, 16–23.
- Heger, Johannes: (Religiöse) Bildung mit KI in der Zukunft?, in: Loccumer Pelikan 1/2025 (i. E.).
- Heger, Johannes: Digital, medial, egal? Religiöse Kompetenz angesichts einer mediatisierten Welt. In: Markus Tomberg und Winfried Verburg (Hg.): RU 4.0. Religiöse Bildung und Digitalisierung, Fulda 2020, 33–60.

- Heger, Johannes: Eschatologie, in WiReLex (2024), abrufbar unter: <https://bibelwissenschaft.de/stichwort/100164/> (abgerufen am 19.06.2024).
- Heger, Johannes: Heilige Lade oder Golem? Hochschul- und (religions-)didaktische Reflexionen zum Umgang mit ChatGPT, in: RU heute 1+2 (2023), 29–34.
- Heger, Johannes: Künstliche Intelligenz – ein Endzeitbeschleuniger?, in: Religion 5-10 57 (2025) (i. E.).
- Heger, Johannes: Künstliche Intelligenz und religiöse Bildung – Vom zweifelnden Bedenken zum bedachten Umgang, in: RelPädplus 4/2024, 36–41.
- Heger, Johannes: Von Halluzinationen und Rationalität. (Praktisch-)Theologische Reflexionen zur Eschatologie im Lichte Künstlicher Intelligenz, in: Karimi, Ahmad Milad (Hg.): Künstliche Intelligenz (= falsafa. Jahrbuch für islamische Religionsphilosophie 7) Baden-Baden 2026 (i. E.).
- Heger, Johannes: Wissenschaftstheorie als Perspektivenfrage?! Eine kritische Diskussion wissenschaftstheoretischer Ansätze der Religionspädagogik, RPG 22, Paderborn 2017.
- Heger, Johannes: Religionsdidaktik als Wissenschaft, in: Kropač, Ulrich/Riegel, Ulrich (Hg.): Handbuch Religionsdidaktik, Stuttgart 2021, 526–536.
- Heinemann, Kerstin: „Die sind nicht ganz so schnell, die ganzen Algorithmen.“ Umgang von Jugendlichen mit Künstlicher Intelligenz und Herausforderungen für pädagogische Settings, in: RU heute 1+2 (2023), 17–22.
- Huwer, Johannes u. a. (Hg.): Kompetenzen für den Unterricht mit und über Künstliche Intelligenz. Perspektiven, Orientierungshilfen und Praxisbeispiele für die Lehramtsausbildung in den Naturwissenschaften, Münster 2024.
- Klar, Maria / Schleiss, Johannes: Künstliche Intelligenz im Kontext von Kompetenzen, Prüfungen und Lehr-Lern-Methoden. Alte und neue Gestaltungsfragen, in: MedienPädagogik 58 (2024), 41–57.
- Kluge, Simon: Künstliche Intelligenz als Thema im Religionsunterricht, in: Jahrbuch der Religionspädagogik 39 (2023), 165–175.
- Knaus, Thomas: Warum KI kein Hype ist und die Medienpädagogik sich damit befassen sollte, in: Medien + Erziehung 68 (2024), 3, 21–30.
- Koerrenz, Ralf / Schlag, Thomas: Editorial, in: Zeitschrift für Pädagogik und Theologie 75 (2023), 2, 125–127.
- Konz, Britta/Ostmeyer, Karl-Heinrich/Scholz, Marcel (Hg.): Gratwanderung Künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI, Stuttgart 2023.
- Köstler, Verena: Zwischen Präzision und Sensitivität: Generierung eines Studienkorpus am Beispiel einer Fragestellung zu Künstlicher Intelligenz (KI) in Bildungsprozessen, in: MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 54 (2023), 1–27.
- Kuka, Lisa / Hörmann, Corinna / Sabitzer, Barbara: Digitaler Dialog, Menschliche Gestaltung – Eine virtuelle Kunstgalerie als Schule der Vernunft: Wie aufgeklärte Medienpädagogik mit AI Literacy verknüpft werden kann, in: Medienimpulse 62 (2024), 1, 1–36.
- Kunz, Sandra: KI-gestützte Lernwelten. Potenziale und Grenzen am Beispiel des Berufsschul-Religionsunterrichts (BRU), München 2024.

- Lindner, Konstantin: Professionalität von Religionslehrer:innen: Anmerkungen aus religionsdidaktischer Perspektive, in: Hailer, Martin / Kubik, Andreas / Otte, Matthias / Schambeck, Mirjam / Schröder, Bernd / Schwier, Helmut (Hg.): Religionslehrer:in im 21. Jahrhundert: Transformationsprozesse in Beruf und theologisch-religionspädagogischer Bildung in Studium, Referendariat und Fortbildung, Leipzig 2023, 61–68.
- Mayrhofer, Florian, Von Jesus-Bots und Luther-Avataren: KI in der religiösen Bildungspraxis, in: *eúangel: Magazin für Missionarische Pastoral* 2 (2024), 1–4.
- Medienpädagogischer Forschungsverbund Südwest (Hg.): JIM 2023. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland, Stuttgart 2023.
- Nassehi, Armin: Gesellschaftliche Grundbegriffe. Ein Glossar der öffentlichen Rede, München 2023.
- Nord, Ilona / Palkowitsch-Kühl, Jens: Nicht die App steht im Mittelpunkt, sondern der Kompetenzerwerb. Ein Praxis-Theorie-Beitrag zur Religionspädagogik in einer mediatisierten Welt, in: *rpi-Impulse* 18 (2018), 3, 5–9.
- Nord, Ilona / Pirner, Manfred: Religiöse Bildung in der digitalen Welt. Die digitale Transformation im Fokus der Religionspädagogik und –didaktik, in: Frederking, Volker / Romeike, Ralf (Hg.): Fachliche Bildung in der digitalen Welt. Digitale Transformation, Big Data und KI im Fokus von 16 Fachdidaktiken, Münster 2022, 330–355.
- Nord, Ilona, Die Bedeutung der Digitalisierung für die religiöse Bildung. Religionspädagogische Diskussionshorizonte und Konkretionen, in: Helmus, Caroline/Riedl, Anna Maria (Hg.): Theologie und Technik. Eine interdisziplinäre Zwischenbilanz (QD 338), Freiburg 2024, 125–148.
- OpenAI: Introducing ChatGPT, in: <https://openai.com/index/chatgpt/> (abgerufen am 26.10.2024).
- Otten, Gabriele / Paeßens, Jutta: Editorial: Künstliche Intelligenz (KI) und Human Enhancement, in: *Religion unterrichten* 2 (2022), 6–7.
- Otten, Gabriele: „Ich bin dein Mensch“: ein humanoider Roboter als Partner und Liebhaber? Gedankenanstöße zur Beurteilung von Formen der Beziehung mit KI Unterrichtsbausteine für die Sekundarstufe II, in: *Religion unterrichten* 2 (2022), 101–120.
- Palkowitsch-Kühl, Jens / Leven, Eva-Maria: Wie begegnen wir künstlicher Intelligenz?, in: *rpi-Impulse* 2 (2020), 18–19.
- Pfaff-Rüdiger, Senta u. a.: Kompass: Künstliche Intelligenz und Kompetenz 2022. Wissen und Handeln im Kontext von KI Bericht zur Repräsentativbefragung im Rahmen des Verbundprojektes Digitales Deutschland (= Digitales Deutschland), München 2022.
- Pirker, Viera: Das Geheimnis im Digitalen. Anthropologie und Ekklesiologie im Zeitalter von Big Data und Künstlicher Intelligenz, in: *Stimmen der Zeit* 144 (2019), 2, 133–141.
- Pirker, Viera: Subjekt Mensch – mehr als ‚das Andere der künstlichen Intelligenz‘? Anthropologie als Grundfrage des Religionsunterrichts (Sekundarstufe I), in: Nord, Ilona / Petzke, Judith (Hg.): Fachdidaktik Religion – diversitätsorientiert und digital. Praxishandbuch für die Sekundarstufe I und II, Berlin 2023, 229–238.
- Platow, Birte: Digitalisierung/Big Data/Künstliche Intelligenz, in: Simojoki, Henrik/Rothgangel, Martin / Körnter, Ulrich H. J. (Hg.): Ethische Kernthemen. Lebensweltlich – theologisch-ethisch – didaktisch, Göttingen 2022, 85–95.

- Platow, Birte: Selbstwahrnehmung und Ich-Konstruktion im Angesicht von Künstlicher Intelligenz, in: Ulshöfer, Gotlind / Kirchschräger, Peter G. / Huppenbauer, Markus (Hg.): Digitalisierung aus theologischer und ethischer Perspektive, Baden-Baden 2021, 105–124.
- Puzio, Anna: Theologie und Künstliche Intelligenz: Perspektiven, Aufgaben und Thesen einer Theologie der Technologisierung, in: Puzio, Anna / Kunkel, Nicole / Klinge, Hendrik (Hg.): Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz (Theologie und Künstliche Intelligenz, Vol. 1), Darmstadt 2023, 13–28.
- ReliBoard: Unterrichtsideen und Tipps, in: <https://reliboard.bistum-essen.de/fachtexte/unterrichtsideen-und-tipps> (abgerufen am 28.10.2024).
- Relilab: Begleitende Forschung, in: <https://relilab.org/forschung/> (abgerufen am 28.10.2024).
- Riegel, Ulrich/Rothgangel, Martin: Religionsdidaktik. Bestandsaufnahmen und Forschungsperspektiven, in: Rothgangel, Martin/Abraham, Ulf/Bayrhuber, Horst u. a. (Hg.): Religionsdidaktik. Bestandsaufnahmen und Forschungsperspektiven aus 17 Fachdidaktiken im Vergleich, Münster. 2. Aufl. (= Fachdidaktische Forschungen, Bd. 13) 2021, 339–362.
- Roebben, Bert: Religionspädagogische Kartographie (= Forum Theologie und Pädagogik, 26), Berlin 2023.
- Rpi-virtuell: Künstliche Intelligenz, in: <https://material.rpi-virtuell.de/schlagwort/kuenstliche-intelligenz/> (abgerufen am 28.10.2024).
- Rysina, Anna / Leven, Ingo: Leben in der digitalen Informationsgesellschaft. Jugendliche und ihr Umgang mit Fakenews und KI, in: Shell Deutschland GmbH (Hg.): Jugend 2024 – 19.Shell Jugendstudie. Pragmatisch zwischen Verdrossenheit und gelebter Vielfalt, Weinheim 2024, 178–183.
- Schlag, Thomas / Yadav, Katharina: Bildungsbeziehungen in der Ankunftszeit des Metaversums – eine religionspädagogische Sensibilisierung, in: Zeitschrift für Pädagogik und Theologie 75 (2023), 2, 163–181.
- Sekretariat der Kultusministerkonferenz (Hg.): Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit künstlicher Intelligenz in schulischen Bildungsprozessen (Beschluss der Kultusministerkonferenz vom 10.10.2024), Berlin 2024.
- Simojoki, Henrik: Digitalisierung – Herausforderung und Chance für den Religionsunterricht. Forschungsbezogene Einordnungen aus religionspädagogischer Perspektive, in: Religionspädagogische Beiträge 82 (2020), 57–63.
- Simone, Meinen: KI im RU, in: <https://digitalrunde.de/2023/03/ki-im-ru> (abgerufen am 28.10.2024).
- Singler, Beth / Watts, Fraser N. (Hg.): The Cambridge companion to religion and artificial intelligence (= Cambridge companions to religion), Cambridge 2024.
- Spiekermann, Sarah: Digitale Ethik und die Künstliche Intelligenz, in: Mainzer, Klaus (Hg.): Philosophisches Handbuch Künstliche Intelligenz (= Springer Reference Geisteswissenschaften), Wiesbaden 2021, 1–24.
- Stalder, Felix: Kultur der Digitalität, Berlin/Grünwald 2016.
- SWK (Hg.): Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz, Bonn 2024.
- Termin, Kathrin: KI–besser als der Mensch? Für künstliche Intelligenz sensibilisieren, in: Grundschule Religion 85 (2023), 28–29.

- Termin, Kathrin: KI-Tools mit Schüler: innen nutzen: Möglichkeiten zum Einsatz von KI im Religionsunterricht, in: *Grundschule Religion* 88 (2024), 28–29.
- Tomberg, Markus / Verburg, Winfried (Hg.): *RU 4.0 Religiöse Bildung und Digitalisierung* (15. Arbeitsforum für Religionspädagogik), Fulda 2020.
- Tulodziecki, Gerhard: Künstliche Intelligenz, in: *Pädagogische Rundschau* 4/2020, 363–378.
- Walldorf, Jochen: Mensch und Maschine. Entwicklungen künstlicher Intelligenz (KI) als Ausgangspunkt anthropologisch-ethischer Lernprozesse in der Oberstufe, in: *rpi-Impulse* 2 (2020), 23–25.
- Wendt, Ruth u. a.: *Algorithmen und Künstliche Intelligenz im Alltag von Jugendlichen. Forschungsbericht für die Bayerische Landeszentrale für neue Medien (BLM) (= BLM-Schriftenreihe – Bayerische Landeszentrale für neue Medien 111)*, Baden-Baden 2024.
- Witt, Claudia de / Gloerfeld, Christina / Wrede, Silke Elisabeth (Hg.): *Künstliche Intelligenz in der Bildung*, Wiesbaden 2023.
- Witt, Theresia / Herrmann, Carmen / Mrohs, Lorenz / Brodel, Hannah / Lindner, Konstantin / Maidanuk, Ilona (Hg.): *Diversität und Digitalität in der Hochschullehre. Innovative Formate in digitalen Bildungskulturen*, Bielefeld 2024.
- Zweig, Katharina A.: *Die KI war's! Von absurd bis tödlich: die Tücken der künstlichen Intelligenz*, München 2023.
- Zweig, Katharina A.: *Künstliche Intelligenz im Religionsunterricht: Wer soll den Menschen richten dürfen?* (2020), in: https://www.erzbistum-koeln.de/export/sites/ebkportal/kultur_und_bildung/-schulen/.content/documentcenter/zeitschrift_impulse/Jahrgang_2020/20imp3basisartikel.pdf (letzter Zugriff: 26.10.2024).

Die Struktur des Digitalen

Lernen über KI an Lernstationen

Michael Köck

Die Erwartungen zum Einsatz Künstlicher Intelligenz im Bildungsbereich sind häufig allein auf die Möglichkeiten gerichtet, Lehr- und Lernprozesse zu effektivieren. Aufgrund des transformativen Potenzials der Technologie für viele Lebensbereiche und vor allem für Arbeit und Beruf ist allerdings auch ein Lernen über KI dringend geboten. Ziel entsprechender Bildungsmaßnahmen muss es sein, den Aufbau von Kenntnissen zur Nutzung und Gestaltung von Anwendungen der Künstlichen Intelligenz zu unterstützen sowie das Beurteilungsvermögen der jeweiligen Zielgruppen in Bezug auf die sozial-humanen Aspekte dieser Technologie zu stärken.

Einen kleinen Beitrag zur Umsetzung dieser Ziele sollen verschiedene Lernstationen leisten, die an der Katholischen Universität Eichstätt-Ingolstadt entwickelt wurden. Vorrangig für die Nutzung in universitären Lehrveranstaltungen im Rahmen der Lehramtsausbildung konzipiert, kommen sie auch bei schulischen und außerschulischen MINT-Maßnahmen zum Einsatz.

Mit den Lernstationen wird nicht der Anspruch erhoben, Künstliche Intelligenz in all ihren Facetten zu erschließen, sondern sich diesem Feld über ihre basalen technologischen Grundlagen anzunähern. Der nachfolgende Beitrag skizziert die didaktischen Vorüberlegungen für die inhaltliche und methodische Konzeption der Lernstationen.

Den Ausgangspunkt hierfür bilden allgemeine Betrachtungen zum Verhältnis von Mensch und Technik, die die Bedeutung des Lernens über digitale Systeme und KI unterstreichen. Der Wandel vom instrumentellen Charakter der Technik hin zu teils autonom agierenden Systemen stellt Individuen und Gesellschaft vor vielfältige Herausforderungen. Wie Erhebungen zu Technikeinstellungen und digitalen Kompetenzen vermuten lassen, scheinen sich viele der Konsequenzen dieser Entwicklungen nicht bewusst zu sein.

Dabei gibt es keinen Mangel an Mahnrufen oder Zielvorstellungen für entsprechende Bildungsmaßnahmen. Jüngstes Beispiel sind die Handlungsempfehlungen der Kultusministerkonferenz (KMK) für den Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen. Gefordert wird hier eine Auseinandersetzung mit Wirkungsweisen, Chancen und Risiken der Künstlichen Intelligenz.

Die Umsetzung solcher Bildungsintentionen stellt ein pädagogisch-didaktisches Problem dar, bei dem konkrete Inhalte ausgewählt und Lernprozesse gestaltet werden müssen. Orientierung bieten hier bewährte didaktische Prozessmodelle.

Sie kamen auch bei der Gestaltung der Lernstationen zum Einsatz. So erfolgte die generelle Auswahl der Lerninhalte anhand von Kriterien wie Exemplarität, Gegenwartsbedeutung und Zukunftsbezug. Die Zuordnung der Inhalte zu den einzelnen Stationen orientiert sich fachlich

an einem Schichtenmodell und methodisch an konkreten arbeitsweltlichen Aufgabenstellungen. Bei der methodischen Konzeption wird auf Prinzipien sowohl des instruktiven als auch handlungsorientierten Lernens gesetzt.

An jeder einzelnen Station lässt sich über eine konkrete Aufgabenstellung ein elementarer Aspekt digitaler Technologien erarbeiten. Im Verbund soll die Bearbeitung, verschiedener, in der Komplexität ansteigender Aufgaben an unterschiedlichen Lernstationen das Verständnis für strukturelle Zusammenhänge fördern.

Erste Evaluationen mit Schülerinnen und Schülern einer 8. Jahrgangsstufe zeigen, dass die Effekte der Arbeit an den Stationen im Rahmen zeitlich begrenzter und isolierter Veranstaltungen nicht unbedingt in konkret nachweisbaren Wissens- oder Fähigkeitszuwächsen, sondern eher in der positiven Wirkung auf einzelne Facetten technikbezogener Selbsteinschätzungen liegen.

Konkrete Leistungszuwächse, etwa im Bereich der informatischen Modellierung realweltlicher Probleme bedürfen daher wohl systematischerer Ansätze in Schule und Unterricht. Nachgewiesen werden konnte im Zuge der Evaluationen, dass die aktuell vorhandenen Stationen, Themen und Aufgabenformate von den Schülerinnen und Schülern interessiert angenommen wurden und vom Schwierigkeitsgrad her angemessen waren. Dies bietet verschiedene Anhaltspunkte sowohl für die Konzeption und Ausgestaltung vergleichbarer MINT-Maßnahmen, als auch für die Weiterentwicklung der Lernstationen.

1 Ausgangslage – Mensch und Technik in Zeiten Künstlicher Intelligenz

Der Digitalisierung und neuerdings der Künstlichen Intelligenz werden Transformationspotenziale zugesprochen wie sonst nur der ersten industriellen Revolution. Damit derartige Aussagen nicht als bloße Floskeln abgetan werden können, bedarf es verschiedener „Reflexionsbewegungen“ über die wechselseitige Beziehung zwischen Mensch und Technik. Entsprechende Deutungsversuche steuern unter anderem Philosophie und Soziologie bei.¹

Mensch und Technik: In ihrem 1960 erschienenen Buch *Vita activa* beispielsweise charakterisiert Hannah Arendt die geschichtliche Entwicklung der Technik aus der Perspektive der Arbeitswelt.² Im vorindustriellen Zeitalter sieht sie die Rollen zwischen Mensch und Technik klar verteilt. Selbst das raffinierteste Handwerkszeug ist hier immer nur Diener seines Herrn, weil selbst unfähig, die Hand zu leiten oder zu ersetzen. Mit der industriellen Revolution und später der „Automation“ hat sich das zunächst rein instrumentelle Verhältnis für die Autorin bereits grundlegend verschoben. Mensch und Welt sind nunmehr dem Rhythmus von Maschinen unterworfen. Wie umfassend sich der Mensch aktuell dem Rhythmus und den Strukturen digitaler Technologien angepasst hat, würde wahrscheinlich auch Hannah Arendt erstaunen. Mehr oder weniger bewusst bewegt sich jeder Einzelne in einer intelligenten,

¹ Nordmann, Alfred: Technikphilosophie, 13.

² Arendt, Hannah: *Vita activa* oder vom tätigen Leben, 134.

digitalisierten und teils autonom agierenden Umgebung. Verbunden damit ist eine Veränderung von Werten, der Wandel von Autoritäten und Institutionen sowie eine Verschiebung gesellschaftlicher und politischer Strukturen.³

Den Anspruch, den Künstliche Intelligenz (KI) dabei erhebt, macht die Definition des BMBF deutlich. Sie beschreibt KI als Teilgebiet der Informatik, bei dem es darum geht, technische Systeme so zu konzipieren, dass sie Probleme eigenständig bearbeiten und sich dabei selbst auf veränderte Bedingungen einstellen können. Solche Systeme haben also die Fähigkeit zu lernen und mit Wahrscheinlichkeiten umzugehen. Damit unterscheiden sie sich von Anwendungen, die klassisch programmiert wurden und damit ihr Leistungsspektrum fest vorgegeben haben.

Bisher hat der technologische Wandel durch digitale Technologien und Künstliche Intelligenz keine größeren Konflikte oder Ängste ausgelöst. Die Deutschen schätzen – so die neuesten Ergebnisse des von der Deutschen Akademie der Technikwissenschaften veröffentlichten Technikradars – die langfristigen Folgen des technischen Fortschritts im europäischen Vergleich sogar überdurchschnittlich optimistisch ein. Gleichwohl hat der Anteil der Skeptiker in den letzten Jahren kontinuierlich zugenommen.

Da ist es umso bemerkenswerter, dass der Wunsch nach bürgerschaftlicher Mitbestimmung in der Technikentwicklung in Teilen der Bevölkerung abgenommen hat. Denn dass sich mit der Künstlichen Intelligenz gerade zentrale Orientierungsgrößen verschieben – zwischen Mensch und Maschine, zwischen Recht und Technik, zwischen Original und Kopie fordert eigentlich zu einer Reflexion der Anwendungsbedingungen und Folgen auf breiter Front heraus.⁴

Verständnis zu Digitalität und KI ausbaubar: Zu Kenntnissen über und Fähigkeiten mit digitalen Technologien existieren einige Untersuchungen. Sie lassen erahnen, dass die durch sie angestoßenen Transformationsprozesse möglicherweise auf ein nur unzureichend ausgeformtes „Navigationssystem“ in der Bevölkerung treffen. Das gilt auch und gerade für Jugendliche – also die Gruppe, der eigentlich eine besondere Nähe zu digitalen Technologien zugebilligt wird.

Die Daten der vor kurzem veröffentlichten ICILS Studie liefern beispielsweise Informationen über die computer- und informationsbezogenen Kompetenzen von Achtklässlerinnen und Achtklässlern in Deutschland.⁵ Erhoben wurden in der Studie u. a. konkrete informatische Kompetenzen. Sie werden auch mit dem Begriff *Computational Thinking* (CT) bezeichnet und gelten als individuelle Fähigkeiten einer Person, realweltliche Probleme informatisch zu modellieren und Lösungsansätze in Zusammenhang mit einem Computer zu entwickeln. Die Studie verdeutlicht, dass nur wenige Jugendliche in diesem Bereich die Leistungsspitze erreichen. Zudem spielen soziale und herkunftsspezifische Faktoren eine entscheidende Rolle.

³ Vgl. Precht, Richard David: Freiheit für alle. Das Ende der Arbeit wie wir sie kannten, 15; vgl. Renn, Jürgen: Die Evolution des Wissens, 279.

⁴ Deutsche Akademie der Technikwissenschaften, acatech: TechnikRadar 2024, 75.

⁵ Eickelmann, Birgit / Fröhlich, Nicole / Bos, Wilfried / Gerick, Julia / Goldhammer, Frank / Schaumburg, Heike / Schwippert, Knut / Senkbeil, Martin / Vahrenhold, Jan (Hg.): *ICILS 2023 #Deutschland*.

Auch über das Ausmaß der Möglichkeiten digitaler Anwendungen und Künstlicher Intelligenz scheinen sich viele Jugendliche nicht im Klaren zu sein. Darauf deuten eigene Erhebungen mit den Schülerinnen und Schülern der achten Jahrgangsstufe eines Gymnasiums in zwei aufeinander folgenden Jahren hin.⁶

2 Bezugsrahmen zum Lernen über KI – grundsätzliche Ziele und Inhalte

Seit digitale Technologien den Verlauf der technologischen Entwicklung prägen, stellt sich die Frage, welche Kenntnisse und Fähigkeiten in allgemeinbildenden Schulen und beruflichen Ausbildungsstätten vermittelt werden sollten. Zusammengefasst werden solche Ziele dann mit Begrifflichkeiten wie „*Information Literacy*“, „informatische Grundbildung“, „Medienkompetenz“, „digitale Kompetenzen“, „*Data Literacy*“ oder „*Computational Thinking*“.

Neuen Schub bekommt der Kompetenzdiskurs nun durch die Künstliche Intelligenz. Sie stellt die Gesellschaft vor die Herausforderung, sich nicht blind den Verlockungen der Digitalisierung zu ergeben. Es gilt vielmehr sinnvolle Einsatzfelder abzustecken und die Chancen und Gefahren klarer zu sehen.⁷

Dies bedingt eine Auseinandersetzung mit grundlegenden Fragen: Für welche Aufgaben sind „intelligente“ Systeme besser geeignet als Menschen? In welchen Situationen sollte das menschliche Urteilsvermögen Vorrang vor maschineller Intelligenz haben? Wie beeinflusst maschinelles Lernen die menschliche Entscheidungsfindung und wo können Maschinen den Menschen in seiner Entscheidungsfindung sinnvoll unterstützen? Wo und warum schleichen sich Fehler und Vorurteile in maschinelle Entscheidungen ein? Wie sehen optimale Schnittstellen zwischen menschlicher und künstlicher Intelligenz aus?⁸

Handlungsempfehlungen der Kultusministerkonferenz (KMK): Die Bedeutung, sich in Schule und Unterricht intensiv mit der Digitalisierung auseinanderzusetzen hat u. a. die Kultusministerkonferenz (KMK) in verschiedenen Papieren herausgestellt. In ihrer erst kürzlich veröffentlichten Handlungsempfehlung zur KI wird betont, dass die Schule die positiven Möglichkeiten der KI mit Blick auf die Förderung von Kindern und Jugendlichen im Lernprozess berücksichtigen und gleichzeitig deutlich machen soll, wie generative KI die Welt verändert.⁹

Gefordert wird in dem Papier der Aufbau eines Verständnisses über die technischen Funktionsweisen sowie über Chancen, Grenzen und Risiken der neuen digitalen Möglichkeiten.

⁶ Köck, Michael: Der Einsatz von Lernstationen zur Förderung des Computational Thinking bei Schülerinnen und Schülern.

⁷ Grunwald, Armin: Der unterlegene Mensch, 242; Precht, Richard David: Künstliche Intelligenz und der Sinn des Lebens, 223f.

⁸ Renn, Jürgen: Die Evolution des Wissens, 782.

⁹ Kultusministerkonferenz (KMK): Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen.

Um den Herausforderungen durch KI verantwortungsvoll zu begegnen, wird die Notwendigkeit des Aufbaus von kritisch-reflexiven Kompetenzen bekräftigt.

Dagstuhl-Dreieck: Den grundsätzlichen Anspruch an eine Bildung in einer digitalen vernetzten Welt verdeutlicht das sogenannte Dagstuhl-Dreieck der Gesellschaft für Informatik aus dem Jahr 2016.¹⁰ Es zeigt, dass es drei gleichberechtigte und sich gegenseitig beeinflussende Kompetenzfelder aufzubauen gilt (s. Abb. 1). Neben Anwendungs- oder Bedienwissen sollen diese Kompetenzen zu Reflexionsprozessen über die persönlichen Nutzungsmöglichkeiten und -gewohnheiten sowie über die gesellschaftlich-kulturellen Auswirkungen digitaler Technologien befähigen.

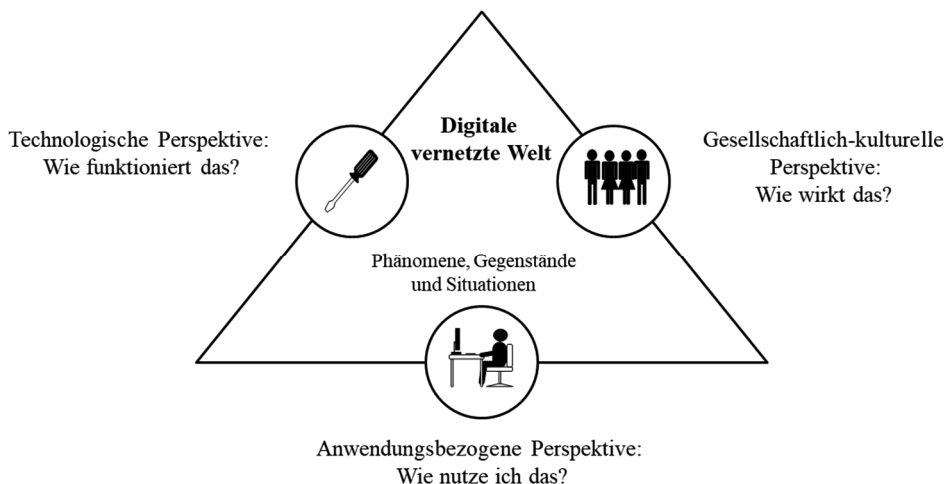


Abbildung 1: „Dagstuhl-Dreieck“ – Gesellschaft für Informatik 2016

Problem der „Bildungsvermittlung“: Die Bildungsbedeutsamkeit eines Wissensbereichs festzustellen ist eine Sache, die Konkretisierung entsprechender Bildungsprozesse eine andere. Die Frage, wie Individuum und Welt, Subjekt und Objekt, Selbst und Anderes aufeinander zubewegt werden können, ist eine zentrale pädagogische Herausforderung.¹¹ Ihre Lösung erfordert sowohl die Auswahl relevanter Inhalte als auch die Entwicklung von Wegen, die den Lernenden vom bloßen Wissen hin zum Verständnis führen. Gerade im Bereich der Künstlichen Intelligenz ist es jedoch nicht einfach, mögliche Lerninhalte zu bestimmen. Dies zeigt auch Künstliche Intelligenz selbst. Laut ChatGPT sollten im Kontext der KI folgende Themen behandelt werden:

- Grundverständnis der Künstlichen Intelligenz
- Anwendungsfelder von KI

¹⁰ Gesellschaft für Informatik e. V.: Dagstuhl-Erklärung. Bildung in der digitalen vernetzten Welt.

¹¹ Derbolav, Josef: Das Selbstverständnis der Erziehungswissenschaft, 128.

- Ethik und KI
- Auswirkungen von KI auf die Arbeitswelt
- KI in der Bildung
- Technikphilosophie und Kritische Theorien zu KI
- Gesellschaftliche Teilhabe und KI
- Zukunftsvisionen und Szenarien

3 Didaktische Fundierung und methodische Konzeption der Lernstationen

Die Vielfalt der als bildungsrelevant eingestuften Inhalte in einen konsistenten didaktisch-methodischen Erklärungszusammenhang zu bringen, ist generell eine anspruchsvolle Aufgabe. Überlegungen, wie sie sich bewältigen lassen, nehmen bis heute ihren Ausgang bei bildungstheoretischen Vorstellungen von Saul B. Robinsohn oder Wolfgang Klafki.

Analyse der Bildungsbedeutsamkeit: Nach Robinsohn sollen Bildung und Erziehung zu einem richtigen und wirksamen Verhalten in der Welt befähigen.¹² Für die Auswahl von Bildungsinhalten skizzierte Robinsohn deshalb drei Kriterien: (1) die Leistung des Bildungsgegenstandes für das Verstehen der Welt, (2) die Bedeutung eines Bildungsgegenstandes für eine sich anschließende Bildungsoption (Berufsausbildung oder Studium) und (3) die Funktion eines Bildungsgegenstandes für bestimmte Verwendungssituationen im privaten und öffentlichen Leben.¹³

Noch näher an die Bildungsrelevanz möglicher Lerninhalte heranzureichen, versprechen die Erschließungsfragen zur didaktischen Analyse von Wolfgang Klafki:¹⁴

- Exemplarität: Welche größeren bzw. allgemeinen Sinn- und Sachzusammenhänge soll ein Inhalt erschließen? Welches Urphänomen oder Grundprinzip, welches Gesetz, Kriterium, Problem, welche Methode, Technik oder Haltung lässt sich in der Auseinandersetzung mit dem Inhalt erfassen?
- Gegenwartsbedeutung: Welche Bedeutung hat der Inhalt beziehungsweise die an diesem Thema zu gewinnende Erfahrung, Erkenntnis, Fähigkeit oder Fertigkeit bereits im geistigen Leben der Lernenden, welche Bedeutung sollte er – vom pädagogischen Gesichtspunkt aus gesehen – darin haben?
- Zukunftsbedeutung: Worin liegt die Bedeutung des Themas für die Zukunft der Lernenden?
- Struktur der Inhalte: Wie lassen sich die Inhalte im Hinblick auf Bedeutungsschichten

¹² Robinsohn, Saul Benjamin: Neue Lern- und Lehrinhalte, 66.

¹³ Robinsohn, Saul Benjamin: Bildungsreform als Revision des Curriculums, 47.

¹⁴ Klafki, Wolfgang: Didaktische Analyse als Kern der Unterrichtsvorbereitung, in: *Die Deutsche Schule*, H. 10, 1958, 450–471.

Klafki, Wolfgang: Didaktische Analyse als Kern der Unterrichtsvorbereitung, in: Roth, Hans / Blumenthal, Alfred (Hg.): *Auswahl Reihe A, Heft 1 – Grundlegende Aufsätze aus der Zeitschrift „Die Deutsche Schule“: Didaktische Analyse*, Hannover 1962, 5–34.

- oder verständnisbildende Zusammenhänge strukturieren, gliedern oder unterteilen?
- Zugänglichkeit: welches sind die besonderen Fälle, Phänomene, Situationen, Versuche, Personen, Ereignisse, Formelemente, in oder an denen die Struktur des jeweiligen Inhaltes den Lernenden dieser Bildungsstufe interessant, fragwürdig zugänglich, begreiflich, anschaulich, werden kann?

Nach wie vor sind diese Fragen bewährte Reflexionshilfen sowohl für die Auswahl von Inhalten als auch für die Strukturierung von Lehr- und Lernprozessen. Sie wurden daher auch für die inhaltliche und methodische Konzeption der Lernstationen eingesetzt.

3.1 Elementare Sinn- und Sachzusammenhänge digitaler Technologien

Abgesehen von Fächern wie Informatik, in denen die Grundlagen digitaler Technologien und Künstlicher Intelligenz systematischer aufgerollt werden können, stellt sich für Lehrkräfte in allen anderen Fächern genauso wie für die Planer und Gestalter einzelner MINT-Maßnahmen in Anlehnung an Klafki eine entscheidende Frage, die auch bei der Konzeption der Lernstationen beantwortet werden musste: Welche elementaren technologischen, ethischen oder auch rechtlichen Sinn- und Sachzusammenhänge sollen und können mit den Lernenden zusammen verständnisbildend erarbeitet werden.

Hilfreiche Anregungen zu dieser Frage kommen vor allem aus der Techniksoziologie, der Technikphilosophie, der Technikdidaktik und natürlich aus der Informatik.

3.1.1 Handlungsspielräume in der digitalen Welt – Vergesellschaftung der Technik oder Technisierung der Gesellschaft

Zur Mündigkeit im Umgang mit Technik gehört immer auch ein Bewusstsein über Technikentstehung und Gestaltung, über Werte und Interessen sowie über Ziele und nicht gewollte Nebenfolgen.¹⁵ Die Anregung entsprechender Reflexionsprozesse stellt eine zentrale Aufgabe einschlägiger Bildungsmaßnahmen dar und ist in Zusammenhang mit digitalen Technologien und Künstlicher Intelligenz von besonderer Bedeutung. Das theoretische Fundament für die Auseinandersetzung mit diesen Themen liefert unter anderem die Techniksoziologie.

Theorien zur Technikentwicklung bieten beispielsweise Zugänge für eine Analyse der Gründe und Konsequenzen des Siegeszugs digitaler Technologien. Die in diesen Theorien formulierten Entscheidungs- und Handlungsspielräume stellen zudem einen Ausgangspunkt für die Erörterung ethischer Fragestellungen dar. Denn um sich der Frage anzunähern, welche Handlungen Menschen im digitalen Zeitalter ausführen dürfen oder sollen, ist es notwendig, das Ausmaß an Entscheidungsfreiheit zu reflektieren, das überhaupt verfügbar ist.

Technikdeterministisch betrachtet ist die Digitalisierung die Folge technischer Sachzwänge, die die gesellschaftlichen genauso wie die individuellen Spielräume beschränken und allenfalls

¹⁵ Grunwald, Armin: Der unterlegene Mensch, 244.

Anpassungsleistungen zulassen.¹⁶ Konstruktivistische Ansätze sehen dagegen eine durchaus aktivere Rolle gesellschaftlicher Kräfte. Denn Entwickler, Konstrukteure oder Entrepreneurinnen sind selbst nicht losgelöst von gesellschaftlichen Verhältnissen und Regelsystemen. Technik ist dann das Ergebnis eines sozialen Aushandlungs- und Aneignungsprozesses.

Beide Ansätze – Vergesellschaftung der Technik und Technisierung der Gesellschaft – können nach Meinung von Günter Ropohl nicht bedenkenlos als Universalansprüche gedeutet werden.¹⁷ Für die Reflexion digitaler Technologien zusammen mit Jugendlichen ergeben sie in jedem Fall zwei Ausgangspunkte für Diskussionen, die auch an den Lernstationen geführt werden sollen.

3.1.2 Struktur und Signatur des Digitalen – Externalisierung, Formalisierung und Materialisierung menschlicher Fähigkeiten

Um in Bildungsmaßnahmen zum Wesen des Digitalen und den elementaren Prinzipien vorzudringen, lohnt eine technikphilosophische Analyse. Von dort aus lässt sich die Struktur des Digitalen ausgehend von geistesgeschichtlichen und technologischen Bedingungen und Entwicklungen beschreiben. Konstitutiv für diese Struktur sind die Prozesse der Externalisierung, Materialisierung, Formalisierung und damit der Abstraktion menschlicher Fähigkeiten auf maschinenrationale Formate:¹⁸

- Menschliche Verstandesfunktionen werden in der digitalen Welt in Form von Symbolen bzw. logischen und mathematischen Operationen externalisiert und mittels elektronischer Schaltkreise materialisiert.
- Die Fähigkeit des Menschen, Reize aus seiner Umgebung über Sinnesorgane aufzunehmen, lässt sich über die Messung physikalischer Größen realisieren und durch Sensorik materialisieren.
- Die eigentlichen Wahrnehmungsfähigkeiten des Menschen sind in Algorithmen bzw. neuronalen Netzen externalisiert und durch leistungsfähige Chips und Prozessoren materialisiert.
- Menschliche Handlungsmöglichkeiten schließlich werden mittels Robotik externalisiert und materialisiert.
- Als grundlegendes Prinzip der Struktur des Digitalen kann die Transformation von Qualitativem in Formal-Quantitatives gelten, das dann die lebensweltliche Seite der Digitalisierung, die Signatur des Digitalen, prägt.¹⁹

Um sich das Ausmaß der technologischen Aneignung menschlicher Fähigkeiten durch digitale Technologien vor Augen zu führen, hilft der analytische Blick auf verschiedene menschliche Handlungsformen. Sie können in das Sprachhandeln, das Bewegungshandeln und in das

¹⁶ Grunwald, Armin: Technikdeterminismus oder Sozialdeterminismus, 64.

¹⁷ Ropohl, Günter: Technologische Aufklärung, 197.

¹⁸ Ebd., 17.

¹⁹ Gramelsberger, Gabriele: Philosophie des Digitalen zur Einführung, 225.

produktbeabsichtigende poetische Handeln unterteilt werden.²⁰ Für alle diese Handlungsformen bieten digitale Technologien und Künstliche Intelligenz bereits Lösungen an.

Kinder und Jugendliche mit der Struktur des Digitalen vertraut zu machen, stellt neben der Reflexion der gesellschaftlich-kulturellen Auswirkungen eine der wichtigsten Aufgaben entsprechender Bildungsmaßnahmen zu digitalen Technologien dar. Der „Überführung“ menschlicher Verstandes- und Handlungsfähigkeiten in digitale Lösungen wird daher an den Lernstationen ein breiter Raum eingeräumt. Hierzu sollen die Lernenden verschiedene arbeits- oder lebensweltliche Probleme zunächst sprachlich formulieren, dann in eine algorithmische Form bringen und diese Abstraktionen anschließend mit graphischen Programmierwerkzeugen in lauffähige Programme für mechatronische Modelle und Systeme umwandeln.

Ethische und rechtliche Problematisierung: Dass Digitalisierung und Künstliche Intelligenz unbestreitbar zu vielen innovativen, effektiven und hilfreichen Anwendungen führen, darf nicht darüber hinwegtäuschen, dass mit dem Programm zur Externalisierung und Materialisierung menschlichen Denkens und Handelns verschiedene Probleme einhergehen. Sie müssen in allen Bildungsprozessen zu digitalen Technologien ausreichend berücksichtigt werden.

Ein drängendes Problem in Zusammenhang mit KI ist zum Beispiel das Verständnisproblem. Bei den heute schon weit verbreiteten Methoden maschinellen Lernens handelt es sich um spezielle Insel-Begabungen. Mit ihnen können Daten zwar hinsichtlich verschiedener Muster und Zusammenhänge interpretiert werden, zu einem Verständnis im menschlichen Sinne führt das jedoch noch nicht. Griffig formuliert, kann hier eher von einem Sieg der Korrelation über die Kausalität gesprochen werden.²¹ Aufgrund der weitreichenden Entscheidungen, zu denen KI-Anwendungen etwa in der Medizin, im Finanzwesen oder beim autonomen Fahren beitragen, ist ein tatsächliches „Verstehen“ der kausalen Zusammenhänge allerdings mehr als geboten.²²

Menschliches Verstehen ist ein komplexer Vorgang, der u. a. Vorerfahrungen voraussetzt. Es ist nicht das Substrat einer Fülle von Daten, sondern speist sich aus individuell erlebten Situationen, bei denen Gefühle, Werte, Zeitbewusstsein und Ich-Identität eine Rolle spielen.²³ Maschinen „verstehen“ anders. Sie geben etwas aus, was ohne die Aggregation großer Datenmengen tatsächlich nicht zu sehen oder zu lernen wäre.²⁴ Das kann positiv sein, jedoch auch negative Folgen zeitigen. Ein Beispiel sind computergenerierte Entscheidungen mit Auswirkungen auf den sozialen Bereich. Gefühle, Stimmungen oder allgemein das Unbestimmte lässt sich aus Daten nur schwer bestimmen, ist aber für menschliches Handeln oftmals unersetzlich.

Die lernfähigen Algorithmen führen auch in anderen Bereichen zu Problemen, etwa im Rechtsbereich. Hier stellt sich beispielsweise das Problem der ungeklärten Zurechnung von

²⁰ Janich, Peter: Kultur und Methode, 34 ff.

²¹ Vgl. Yogeshwar, Ranga: Nächste Ausfahrt Zukunft, 191

²² Brandstetter, Thomas: Nicht ohne Grund!, 24 ff.

²³ Precht, Richard David: Künstliche Intelligenz und der Sinn des Lebens, 22 ff.

²⁴ Vgl. Bedorf, Thomas: Maschinenhermeneutik, 22.

Verantwortlichkeiten. Denn überall dort, wo KI-Systeme zum Einsatz kommen, stößt die Anknüpfung rechtlicher Folgen an willentliches Verhalten an Grenzen. Beispiele dafür sind etwa selbstlernende Rasenmäher, die erst auf dem Grundstück des Käufers gelernt haben, dass kleine weiche Gegenstände unproblematisch überfahren werden können und den Schoßhund der Nachbarin dabei töten oder ein Algorithmus zur Preisgestaltung im Internet, der eigenmächtig einen Preiskorridor verlässt und dem Händler empfindliche Verluste einbringt. In beiden Fällen zeigt sich eine Haftungs- und Verantwortungslücke. Sie entsteht durch die problematische Zurechnung der Handlungen des KI-Systems jeweils zum Produzenten oder Anwender.²⁵ Juristisch ist dieses Problem noch nicht gelöst – als Anlass für Diskussionen zu KI aber durchaus geeignet.

3.1.3 Methoden und Konzepte der KI – Künstliche Entsprechungen menschlicher Intelligenz

Anwendungsbereiche und Begrifflichkeiten im Kontext künstlicher Intelligenz nehmen von Tag zu Tag zu. Hier den Überblick zu behalten, fällt daher schwer. Um diesem Problem zu begegnen, hat der Computerwissenschaftler Kris Hammond ein „Periodensystem der Künstlichen Intelligenz“ entworfen. Jedes der 28 KI-Element repräsentiert eine Teilfunktion. Aus der Kombination von drei Elementen ergibt sich eine KI-Anwendung.²⁶ Hinter jedem dieser Elemente stehen wiederum unterschiedliche Methoden, Techniken und Algorithmen. Ihre genaue Kenntnis muss Informatikern vorbehalten bleiben. Um jedoch zu verstehen, warum überhaupt von Künstlicher Intelligenz gesprochen wird, lohnt es sich, in entsprechenden Bildungsprozessen Vergleiche zu menschlichen Denkvorgängen herzustellen.

Informatische Entsprechungen menschlicher Denkvorgänge: Menschliche Intelligenz wird gemeinhin als Summe unterschiedlicher Denk- und Wahrnehmungsprozesse gesehen. Diese Prozesse können sich vom Anforderungsniveau her deutlich unterscheiden. Dies gilt auch für die Künstliche Intelligenz. Während in manchen Softwareanwendungen die Intelligenz mit klaren wenn-dann-Regeln bereits „hineincodiert“ wurde, sind andere Systeme zu induktiven Schlüssen oder sogar zum Modelllernen fähig. Das bedeutet, dass es nicht „die“ oder „eine“ KI gibt, sondern unterschiedliche Konzepte und Methoden, um mit bestehenden Daten und Informationen umzugehen oder neue zu kreieren. Didaktisch vereinfacht lassen sich diese Methoden oder Techniken als Entsprechungen menschlicher Verstandesfähigkeiten charakterisieren (s. Abb. 2).

Die Reproduktion von Wissen entlang bestimmter Regeln beispielsweise ist ein menschliches Verhalten, das sich mit dem regelbasierten Abruf von Informationen aus Datenbanken vergleichen lässt. Auch das menschliche Schlussfolgern kann algorithmisch nachgebildet werden, und zwar über sogenannte semantische Netze. Dabei wird ein Modell mit Begriffen und Beziehungen strukturiert. Das erlaubt die Ableitung von Beziehungen, Zusammenhängen

²⁵ Riehm, Tobias: *Roboter als Personen im Rechtssinne?*, 30 ff.

²⁶ Bitkom e.V.: *Digitalisierung gestalten mit dem Periodensystem der Künstlichen*, 15.

und Bedeutungen. Mit unterschiedlichen statistischen Verfahren wiederum können Muster in Daten ausfindig und auf diese Weise Zusammenhänge identifiziert werden. Und mit speziellen algorithmischen Strukturen wie Entscheidungsbäumen oder neuronalen Netzen können Maschinen sogar „Lernen“.

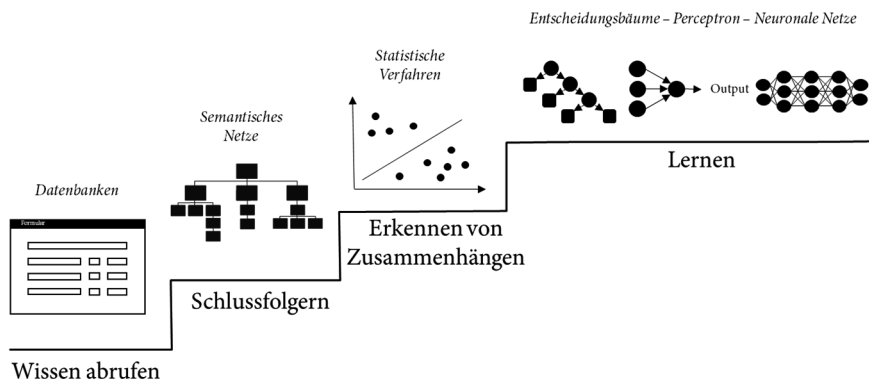


Abbildung 2: Techniken, Methoden und Algorithmen der Künstlichen Intelligenz – Michael Köck

Umkehrung des Programmierparadigmas: Während bei klassischen Programmen Regeln in Programmcode übersetzt werden, mit denen sich anschließend Daten auswerten lassen, erfolgt bei statistischen Ansätzen der KI eine Umkehrung. Hier werden entweder Daten vorgegeben und mit statistischen Werkzeugen Muster gesucht oder aber Daten und Antworten vorgegeben. In diesem Fall werden Modelle erstellt, auf deren Grundlage dann Vorhersagen für andere Daten getroffen werden.

Im Rahmen einzelner MINT-Maßnahmen können die verschiedenen Methoden oder Techniken der KI allenfalls erwähnt, aber nur schwer eingehender behandelt werden. Dass es im Verlauf der „Intelligenzstufen“ allerdings zu einer Umkehrung des klassischen Programmierparadigmas kommt, ist ein Charakteristikum der KI, das bei entsprechenden Veranstaltungen Berücksichtigung finden sollte.

Um diesen Unterschied zwischen den Techniken der KI im Rahmen der Lernstationen deutlich zu machen, sind für einige Aufgaben an den Lernstationen Algorithmen mit klaren Regeln zu erstellen, an anderen dagegen zunächst Modelle mit Daten zu trainieren.

Datenschutz und Diskriminierung: Mit der Menge an erforderlichen Daten (Big Data), die für KI-Anwendungen notwendig ist, sind einige Probleme verbunden. Einige davon lassen sich auch in Zusammenhang mit den Lernstationen thematisieren. Dies betrifft beispielsweise das Problem personenbezogener Daten. Denn überall dort, wo sie erhoben und verarbeitet werden, kann dies zu rechtlich fragwürdigen Konsequenzen führen (vgl. Kunkel & Schoewe

2021, 9ff.).²⁷ Ein Problembereich ist etwa das Profiling, also das Sammeln personenbezogener Daten zum Zweck der Kategorisierung.

Dass Algorithmen auf Topologien, Kategorien und Hierarchien basieren, die über Variablen und Methoden festgelegt werden, führt zu weiteren Schwierigkeiten. Denn mit ihnen sind bestimmte Normen, Werte und Entscheidungen der Entwickler festgeschrieben. Welche Voraussetzungen oder Kriterien bei ihrer Auswahl eine Rolle gespielt haben, bleibt oft intransparent. Dadurch kann es zu Benachteiligungen von Menschen kommen – etwa wenn spezielle Daten über bestimmte Personengruppen nicht berücksichtigt werden oder Trainingsdaten einseitig auf spezielle Personengruppen ausgerichtet sind. In Verbindung mit den personenbezogenen Daten besteht somit die Gefahr von Diskriminierung sozialer Gruppen. Als Beispiel lässt sich ein Bewerbungsverfahren anführen.²⁸

Moralisch-ethische Probleme in Zusammenhang mit der Interpretation von Daten werfen nicht nur trainierte Systeme auf. Denn mittlerweile haben KI-Systeme den Beweis erbracht, dass sie auf der Basis einer bestimmten Anzahl an Regeln und Lernvorgängen selbst Entscheidungsstrategien generieren können. Eine der drängenden Fragen der Zukunft wird daher sein, ob dem Ausmaß an Vertrauen in die Entscheidungen von Maschinen ein ebenso großes Verständnis ihrer Fehlbarkeit gegenübersteht.

3.2 Gegenwarts- und Zukunftsbedeutung

Digitalisierung meint die Umwandlung bisher analoger Prozesse und Produkte in digitale. In Abgrenzung dazu beschreibt Digitalität dann den jeweiligen Zustand, der durch Digitalisierung entstanden ist. Ein Kennzeichen dieses Zustandes ist nach Felix Stalder die Algorithmizität.²⁹ Damit meint er u. a. die Determinierung menschlicher Handlungen durch Algorithmen.

Algorithmen steuern entweder direkt Systeme, Maschinen und Abläufe oder sie werden für die Analyse und Verarbeitung von Daten eingesetzt. Dabei verändern sie die damit in Verbindung stehenden Handlungszusammenhänge oder schaffen ganz neue. Das gilt natürlich auch für die Welt von Kindern und Jugendlichen. Die digitalen Auswirkungen auf Lebensbereiche wie Freizeit und Konsum, Arbeit und Beruf sowie politische Partizipation sind bereits jetzt umfassend und nichts deutet darauf hin, dass sich das umkehrt. Nachfolgend einige Beispiele, anhand derer sich die Bedeutung von Kenntnissen und Kompetenzen in Zusammenhang mit den digitalen Grundlagen der KI begründen lassen.

Auswirkungen digitaler Technologien auf Konsum: Kinder und Jugendliche sind Nutzende von Informations-, Kommunikations- und Netzwerktechnologien. Als bereits zahlungs-

²⁷ Vgl. Kunkel, Christian / Schoewe, Johannes: Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall, 9ff.

²⁸ Vgl. Lemke, Christoph / Monett, David / Bloomfield, Michael: Lernen und lehren mit und über KI, 55 ff.

²⁹ Stalder, Felix: Kultur der Digitalität.

kräftige Zielgruppe werden sie mit den verschiedensten digitalen Werbeformaten gezielt umworben. Sie kaufen, bezahlen und bewerten über das Internet und sind dabei oftmals unbewusst in der Rolle von Anbietenden, etwa durch die Weitergabe persönlicher Informationen und Daten.³⁰

Dieser bereits vorsortierte „Datenrohstoff“ lässt sich u. a. mit Anwendungen der künstlichen Intelligenz verarbeiten, interpretieren und dann weiter gewinnbringend einsetzen. Notwendig ist daher ein Bewusstsein über die ökonomischen Verwertungsinteressen, die ausschlaggebend sind für die Entwicklung digitaler Technologien und die Sammlung von Verbraucherdaten. Zusätzlich benötigen Jugendliche Kenntnisse, auf welche Weise personenbezogene Daten gewonnen und gespeichert sowie für die Einflussnahme genutzt werden können.

Auswirkungen digitaler Technologien auf Arbeitswelt und Berufe: Ein wichtiges Ziel sowohl der allgemeinbildenden Schulen als auch zahlreicher MINT-Maßnahmen ist die Förderung der Berufswahlkompetenz von Jugendlichen. Diese beruht neben einem entwickelten vorberuflichen Selbstkonzept auf ausreichend Kenntnissen über arbeitsweltliche Strukturen und Berufe. Für die Berufs- und Studienorientierung ist die Digitalisierung daher Auftrag, Schlagworte mit Inhalten zu füllen und handlungsorientierte Zugänge zu beruflichen Anforderungen in einer digitalisierten Arbeitswelt zu ermöglichen.

Digitalisierung bedeutet grundsätzlich eine Umwandlung bisher analoger in digitale Prozesse und Produkte mit Auswirkungen auf Märkte, Branchen und die dort tradierten Arbeitsstrukturen. Potenziell kann die Digitalisierung zur Substitution menschlicher Arbeit durch Maschinen, zu effektiven Formen der Assistenz für den arbeitenden Menschen und zur Entgrenzung und Flexibilisierung der Arbeitsorte und -zeiten führen.

Im beruflichen Bereich führt die Digitalisierung im Einzelfall zur Reduktion, Transformation, Modifikation oder Einrichtung neuer Berufe. Generell ändern sich durch digitale Technologien in nahezu allen Berufen die Arbeitsbedingungen, Aufgabenzuschnitte und Anforderungen. Insbesondere Anwendungen der Künstlichen Intelligenz schicken sich an, in Bereiche vorzudringen, in denen noch bis vor kurzem menschliche Fähigkeiten als unersetzbar galten. Dazu zählen zum Beispiel kognitive und manuelle Flexibilität, der Umgang mit variablen Situationen oder die Abstraktions-, Anpassungs- und Lernfähigkeit.

In der öffentlichen Diskussion werden oft allein die Rationalisierungspotenziale der Künstlichen Intelligenz in den Vordergrund geschoben. Dabei können KI-Systeme prinzipiell auf zwei Wegen in Arbeitsprozesse integriert werden: Entweder substituierend oder komplementär zu menschlicher Arbeit. Entsprechend eingesetzt, liefern sie zudem einen Beitrag zur Humanisierung der Arbeitswelt, verringern eine Dequalifizierung und können zu lernförderlichen Arbeitsbedingungen führen.³¹

³⁰ Schauer, Christian / Frank, Ulrich: Wirtschaftsinformatik an Schulen, 1.

³¹ Becker Matthias / Spöttl, Georg / Windelband Lars: Digitale Transformation, 7.

Dass die großen Datenmengen, auf denen Künstliche Intelligenz aufbaut, oftmals unter prekären Arbeitsbedingungen generiert werden, ist ein weiterer Punkt, den es beim Blick auf den Zusammenhang zwischen KI und Arbeitswelt zu beachten gilt.

In welchem Ausmaß sich der Arbeitsmarkt unter dem Einfluss von KI verändert, kann derzeit kaum sicher prognostiziert werden – zu rasant ist die Entwicklung. Mit Sicherheit sind jedoch massive strukturelle Veränderungen zu erwarten. Ob die positiven Aspekte die Risiken übertreffen, hängt nicht zuletzt von politischen Rahmenbedingungen ab.

Auswirkung digitaler Technologien auf Politik: Für die politischen Rahmenbedingungen stellen digitale Technologien und Künstliche Intelligenz allerdings selbst große Herausforderungen dar. Wie Yuval Noah Harari befürchtet, laufen demokratische Institutionen wie Wahlen, Parteien oder Parlamente Gefahr, weder die Datenmengen noch deren Umlaufgeschwindigkeit angemessen zu bewältigen.³² Für Armin Grundwald stehen Demokratie und Digitalisierung ohnehin in einem grundsätzlichen Gegensatz. Denn während Demokratie auf das mühsame Aushandeln komplexer Fragen und anstrengende Aushandeln verschiedener Interessen setzt, bietet die Automatisierung scheinbar schnelle und optimale Lösungen.³³

KI-Anwendungen sollten deswegen auch aus gesellschaftlicher und politischer Sicht aufmerksam begleitet und kontrolliert werden. Denn die ihnen eingeschriebenen und wenig transparenten Werte und Normen verbreiten sich durch die Nutzung in der Gesellschaft und haben das Potenzial, Personengruppen zu diskriminieren, die Informations- und Meinungsfreiheit zu beschränken oder Einstellungen und Identitäten zu formen.

3.3 Bedeutungsschichten und verständnisbildende Zusammenhänge

Wie die vorausgehenden Analysen zeigen, umfasst das Lernen über Künstliche Intelligenz, wirtschaftliche, technologische, informatische, ethische und rechtliche Aspekte. Eine grundsätzliche Frage war daher, wie diesen Punkten an den einzelnen Lernstationen ausreichend Geltung verschafft werden kann und wie sie gleichzeitig inhaltlich so zu strukturieren sind, dass sich verständnisbildende Zusammenhänge in Bezug auf technologische und human-soziale Aspekte ergeben können.

Eine schlüssige Zuordnung von Inhalten zu einzelnen Lernstationen ergab sich dabei durch eine Orientierung an Ebenen- bzw. Schichtmodellen, wie sie innerhalb der Informatik Verwendung finden. Derartige Modelle helfen bei der Klärung der Architektur informations- und kommunikationstechnischer Systeme und Anwendungen. Sie bilden zudem die Grundlage für die domänenspezifische Kompetenz in der Informatik – nämlich Probleme zu analysieren, also zu zerlegen und sie auf das Wesentliche zu reduzieren.

³² Harari, Yuval Noah: *Homo deus*, 506.

³³ Grundwald, Armin: *Der unterlegene Mensch*, 170 f.

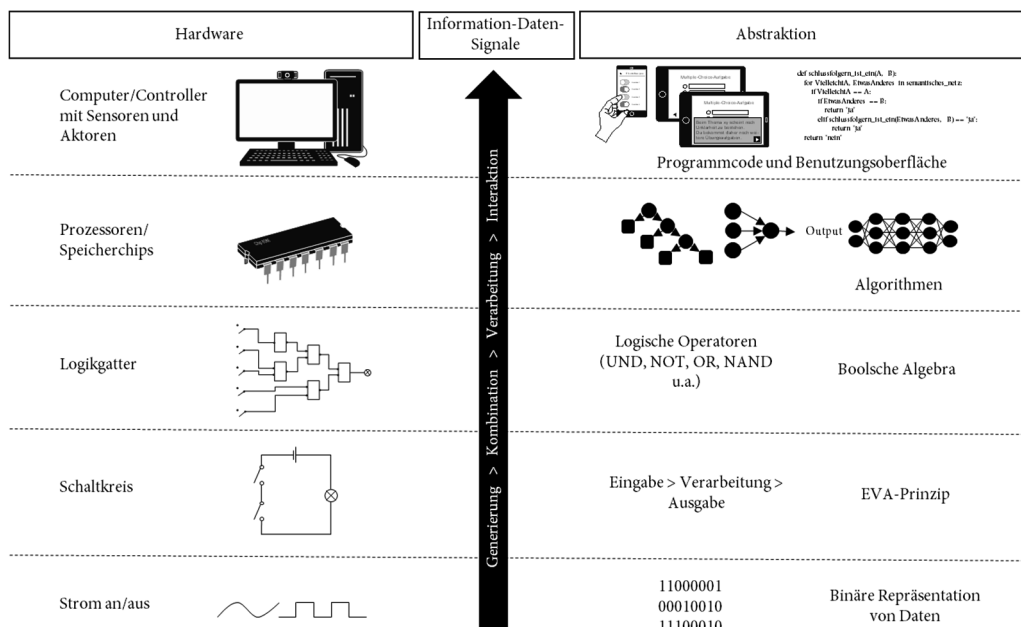


Abbildung 3: Einfaches Schichtensystem informationstechnischer Zusammenhänge – Michael Köck

Ein solches Schichtenmodell lässt sich bilden, indem auf einer Ebene jeweils die Abstraktionen sprachlich formulierbarer menschlicher Denkvorgänge und ihrer Materialisierungen in Form elektronischer Systeme zusammengefasst werden. Auf diese Weise können verschiedene Ebenen definiert werden, die an inhaltlicher Komplexität zunehmen. Sie reichen von den elementaren Grundlagen digitaler Technologien, wie zum Beispiel dem Binärsystem, bis hin zu komplexen Systemen und Anwendungen (s. Abb. 3).

Die Arbeit an den einzelnen Lernstationen zielt dabei jeweils auf ein grundlegendes Verständnis ausgewählter „Schichten“ mit ihren technologischen Charakteristika. Die thematische Einbindung der Technik in ein lebensweltliches Ausgangsproblem, das in Form einer Aufgabe zu lösen ist, soll zugleich eine Reflexion sozial-humaner Implikationen digitaler Technologien ermöglichen. Schwerpunkte der Aufgaben bilden dabei die Bereiche Arbeit und Beruf.

3.4 Zugänglichkeit – Methodische Strukturierung

Zielsetzungen des methodischen Settings: Aus methodischer Sicht stellten sich in Bezug auf die Lernstationen vor allem die folgenden beiden Fragen:

Wie kann ein dauerhaftes, didaktisch auf verschiedene Zielgruppen adaptierbares, curricular verankertes sowie anschluss- und erweiterungsfähiges Bildungsangebot konzipiert

werden, das im universitären, schulischen und außerschulischen Bereich eingesetzt werden kann? Wie muss eine Lernstrecke konzipiert sein, die Lernenden eine möglichst selbsttätige Erschließung komplexer Lerngegenstände erlaubt und den Kompetenzzuwachs durch bereits Gelungenes, Verstandenes und Vollendetes spürbar werden lässt?

Die Antwort auf diese Fragen wurde in einem erweiterbaren Set an Lernstationen gesehen, die sich sowohl für Phasen der Instruktion als für eine aktive, selbstgesteuerte Aneignung von Wissen und Problemlösungsmöglichkeiten eignen.

Aufgabenorientierung der Lernstationen: Thematisch lassen sich die Lernstationen grob einzelnen Ebenen des vorher skizzierten Schichtenmodells zuordnen. Sie unterscheiden sich durch das Ausmaß an Komplexität sowohl in Bezug auf die Abstraktionen als auch in Bezug auf die technischen Entsprechungen (s. Tab. 1).

Erkenntnis und Erfahrung sollen sich an den Stationen möglichst an konkreten Aufgaben entwickeln. So kann das Dualsystem an einem „Binärzähler“ entweder durch forschendes Lernen oder durch konkrete Rechenaufgaben erarbeitet werden. Am Modell eines Parkhauses soll die Belegung durch Fahrzeuge kenntlich gemacht werden. Dabei lässt sich die Bedeutung elektrischer Signale für die Informationsübermittlung erfahren. Mit verschiedenen logischen Schaltungen bzw. Gattern können die Möglichkeiten der Veränderung von Informationszuständen untersucht und auf konkrete Aufgabenstellungen übertragen werden. Bei der Lösung von Aufgaben, bei denen Controller und Roboter programmiert werden müssen, setzen sich die Lernenden mit den hard- und softwaretechnischen Grundlagen mechatronischer Systeme auseinander und programmieren mit graphischen Werkzeugen klassische Algorithmen. Aufgaben mit einem 3D-Drucker ermöglichen den Lernenden Einblicke in die Produktentwicklung und -erstellung.

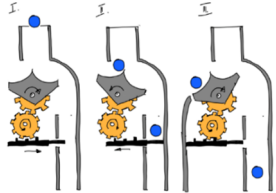
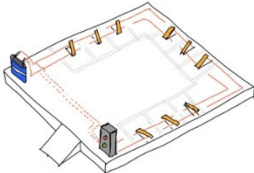
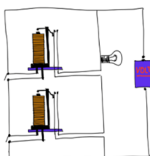
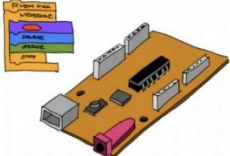
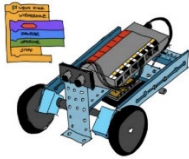
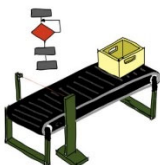
Drei Stationen sollen gezielt Zugänge zur Beurteilung von Chancen und Risiken der Künstlichen Intelligenz schaffen. Mit Hilfe des „analogen“ NIM-Spiels gewinnen die Lernenden zunächst einen Einblick in die grundsätzliche Funktionsweise des maschinellen Lernens. Sie erfahren, wie durch Training Regeln abgeleitet oder auch komplexe Muster erkannt werden können.

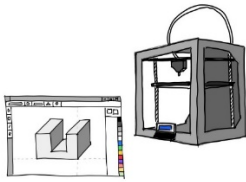
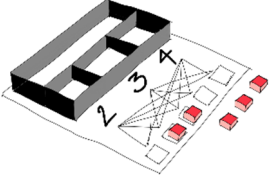
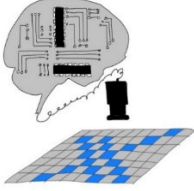
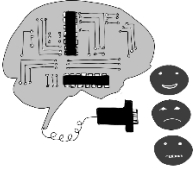

Die Bedeutung von Trainingsdaten für die Künstliche Intelligenz lässt sich außerdem an einer Station erkennen, in der es um Qualitätssicherungsprozesse im Rahmen der Produkterstellung geht. Ebenfalls trainiert werden muss die Datengrundlage für ein Gesichtserkennungssystem, das als Türöffner fungiert. Zusätzlich kann an einer Station mit einem Roboterarm, der sich durch Handführung programmieren lässt, die Reichweite digitaler Technologien im industriellen Bereich veranschaulicht werden.

Ausstattung der Lernstationen: Für die Ausstattung der Lernstationen wurde auf eine Kombination unterschiedlicher Elemente gesetzt. Mit Ausnahme des 3D-Druckers und des Roboterarms werden relativ preiswerte Gerätschaften eingesetzt. Erforderliche Aufbauten und Konstruktionen wurden teilweise selbst erstellt, sodass sie sich auch zur Nachahmung für studentische oder schulische Projekte eignen. Softwaretechnisch kommen die graphischen Programmierumgebungen Scratch oder mBlock zum Einsatz, wobei sich einige der Stationen auch für eine Anwendung textorientierter Sprachen eignen.

Für die Unterbringung und Präsentation der Lernstationen wurden fahr- und verschließbare Kästen angeschafft. Sie sollen eine flexible Anordnung im Raum und einen leichteren Transport der Modelle und Gerätschaften für den außeruniversitären Einsatz ermöglichen. Glasaufsätze für die Kästen bieten sich zusätzlich für themenbezogene Ausstellungen an.

Tabelle 1: Übersicht über die Lernstationen

<p>1 Station Binärzähler: Grundlage für die Digitalisierung – das Zählen mit nur zwei Ziffern</p>	<p>2 Station Parkhaus: Stromkreis als Voraussetzung für den Signal- und Datentransport</p>
	
<p>3 Station Elektronische Schaltungen: Neue Informationszustände schaffen durch Schaltungen</p>	<p>4 Station Arduino: Mikrocomputer – Wenn Technik reagiert, braucht es ein „Gehirn“</p>
	
<p>5 Station Roboter: Programmieren mit Scratch – wie kann Robotern Bewegung „einghaucht“ werden?</p>	<p>6 Station Förderband: Maschinen in der Arbeitswelt – wie komplexe technische Prozesse gesteuert werden</p>
	
<p>7 Station 3D-Drucker: Vom Konsumenten zum Produzenten</p>	<p>8 Station Künstliche Intelligenz NIM-Spiel: Lernen durch Training</p>

	
<p>9 Station Künstliche Intelligenz: Trainierte Entscheidungen zur Qualitätssicherung</p>	<p>10 Station Künstliche Intelligenz: Gesichtserkennung mit KI</p>
	
<p>11 Station Kollege Roboter: Programmierte Verantwortung</p>	
	

Zu einzelnen Lernstationen existiert ein umfangreiches Angebot an Informationen im Internet in Form eines Ilias-Kurses.³⁴ Dort werden zu jeder Station Ziele, erforderliches Fachwissen, Überlegungen zur Bildungsrelevanz sowie methodische Vorschläge, Anleitungen und Arbeitsmaterialien präsentiert, teilweise ergänzt um Filme und simulative Apps.

Die Materialien übernehmen verschiedene Funktionen: Sie dienen als Vorbereitung für die Arbeit an den Lernstationen sowohl für Lehrende als auch Lernende, bieten Studierenden fachliche und didaktische Zugänge für eigene Projekte und sollen außerdem sicherstellen, dass Stationen und Materialien über den an Hochschulen üblichen Personalwechsel hinaus in Gebrauch bleiben.

³⁴ https://elearn.ku.de/goto.php?target=crs_731417&client_id=elearnKU (abgerufen am 12.09.2025).

5 Reflexion

Die Lernstationen kamen u. a. in zwei aufeinanderfolgenden Jahren bei Veranstaltungen mit Schülerinnen und Schülern der achten Jahrgangsstufe eines Gymnasiums zum Einsatz. Die Veranstaltungen wurden mithilfe eines Prä-Post-Designs evaluiert. Auf diese Weise sollten Erkenntnisse sowohl zu den Effekten der Arbeit mit den Lernstationen als auch zur Weiterentwicklung des Angebots an Lernstationen gewonnen werden.³⁵

Das vorrangige Ziel der Untersuchungen war, die Bandbreite möglicher Formate und Schwierigkeitsgrade von Aufgaben auszuloten, die einen für Schülerinnen und Schüler motivierenden Lernprozess erwarten lassen. Ein weiteres Ziel war es, Erkenntnisse darüber zu erlangen, ob das Veranstaltungsformat einen Beitrag zum Verständnisaufbau für Digitalisierungsprozesse speziell in der Arbeitswelt leisten kann und möglicherweise sogar zu einem Zuwachs an technologie- bzw. computerbezogener Problemlösekompetenz führt.

Die erhobenen Daten zeigen, dass die Effekte der Arbeit an den Stationen im Rahmen zeitlich begrenzter und isolierter Veranstaltungen nicht unbedingt in konkret nachweisbaren Wissens- oder Fähigkeitszuwächsen zu suchen sind, sondern eher in der positiven Wirkung auf einzelne Facetten technikbezogener Selbsteinschätzungen.

Konkrete Leistungszuwächse, etwa im Bereich der informatischen Modellierung realweltlicher Probleme, bedürfen systematischerer Ansätze in Schule und Unterricht. Allerdings lässt die Evaluation die Schlussfolgerung zu, dass die aktuell vorhandenen Stationen, Themen und Aufgabenformate von den Schülergruppen interessiert angenommen wurden und vom Schwierigkeitsgrad her angemessen waren.

Dies bietet verschiedene Anhaltspunkte sowohl für die Konzeption und Ausgestaltung vergleichbarer MINT-Maßnahmen, als auch für die Weiterentwicklung der Lernstationen.

Literatur

Arendt, Hannah: *Vita activa oder vom tätigen Leben*. 6. Auflage, München, Zürich 1989.

Bedorf, Thomas: Maschinenhermeneutik, in: Gerlek, Sebastian / Kissler, Sarah / Mämecke, Till / Möbus, Daniel (Hg.): *Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen*, Hagen 2022, 16–31.

Becker Matthias/ Spöttl, Georg/ Windelband Lars: Digitale Transformation. Digitalisierte Arbeitswelt und Konsequenzen für die berufliche Erstausbildung. *Bildung und Beruf*. 6/2023, 6–15.

Bitkom e.V.: Digitalisierung gestalten mit dem Periodensystem der Künstlichen Intelligenz – Ein Navigationssystem für Entscheider, online verfügbar unter: https://www.bitkom.org/sites/main/files/2018-12/181204_LF_Periodensystem_online.pdf, 2018 (abgerufen am 10.12.2024).

Brandstetter, Thomas: Nicht ohne Grund! In: *Max Planck Forschung* 3/2021, 21–31.

³⁵ Köck, Michael: Der Einsatz von Lernstationen zur Förderung des Computational Thinking.

- Deutsche Akademie der Technikwissenschaften, acatech: *TechnikRadar 2024. Was die Deutschen über Technik denken*, online verfügbar unter: <https://www.acatech.de/publikation/technikradar-2024/>, 2024 (abgerufen am 10.12.2024).
- Derbolav, Josef: Das Selbstverständnis der Erziehungswissenschaft, in: Oppolzer, Sigbert (Hg.): *Ausgangspunkte wissenschaftlichen Denkens: Denkformen und Forschungsmethoden der Erziehungswissenschaft. Band 1: Hermeneutik Phänomenologie Dialektik Methodenkritik*, München 1966, 119–158.
- Eickelmann, Birgit / Fröhlich, Nicole / Bos, Wilfried / Gerick, Julia / Goldhammer, Frank / Schaumburg, Heike / Schwippert, Knut / Senkbeil, Martin / Vahrenhold, Jan (Hg.): *ICILS 2023 #Deutschland. Computer- und informationsbezogene Kompetenzen und Kompetenzen im Bereich Computational Thinking von Schülerinnen im internationalen Vergleich,** Münster 2024, online verfügbar unter: <https://doi.org/10.31244/9783830999492>.
- Gesellschaft für Informatik e. V.: Dagstuhl-Erklärung. Bildung in der digitalen vernetzten Welt, online verfügbar unter: https://gi.de/fileadmin/GI/Hauptseite/Themen/Dagstuhl-Erklärung_2016-03-23.pdf, 2016 (abgerufen am 10.12.2024).
- Gramelsberger, Gabriele: *Philosophie des Digitalen zur Einführung*, Hamburg 2023.
- Grunwald, Armin: *Der unterlegene Mensch. Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern*, München 2019.
- Grunwald, Armin: Technikdeterminismus oder Sozialdeterminismus: Zeitbezüge und Kausalverhältnisse aus der Sicht des „Technology Assessment“, in: Dolata, Ulrich / Werle, Raymund (Hg.): *Gesellschaft und die Macht der Technik: Sozioökonomischer und institutioneller Wandel durch Technisierung*, Frankfurt/Main 2007, 63–82.
- Harari, Yuval Noah: *Homo deus. Eine Geschichte von Morgen*, Sonderausgabe für die Bundeszentrale für politische Bildung, Bonn, 2017.
- Janich, Peter: *Kultur und Methode. Philosophie in einer wissenschaftlich geprägten Welt*, Frankfurt am Main 2006.
- Köck, Michael: Der Einsatz von Lernstationen zur Förderung des Computational Thinking bei Schülerinnen und Schülern, in: *Journal of technical education: JOTED* 12 (29. April 2024), 2, 101–128.
- Klafki, Wolfgang: Didaktische Analyse als Kern der Unterrichtsvorbereitung, in: *Die Deutsche Schule*, H. 10, 1958, 450–471.
- Klafki, Wolfgang: Didaktische Analyse als Kern der Unterrichtsvorbereitung, in: Roth, Hans / Blumenthal, Alfred (Hg.): *Auswahl Reihe A, Heft 1 – Grundlegende Aufsätze aus der Zeitschrift „Die Deutsche Schule“: Didaktische Analyse*, Hannover 1962, 5–34.
- Kultusministerkonferenz (KMK): Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen, online verfügbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pdf, 2024.
- Kunkel, Christian / Schoewe, Johannes: Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall einschließlich Profiling im Sinne Art. 22 DSGVO – Praxisrelevanz und Wirksamkeit der Norm in Zeiten von Big Data und KI, in: Barton, Thomas / Müller, Claudia (Hg.): *Künstliche Intelligenz in der Anwendung. Rechtliche Aspekte, Anwendungspotenziale und Einsatzszenarien*, Wiesbaden 2021, 9–23.

- Lemke, Christoph / Monett, David / Bloomfield, Michael: Lernen und lehren mit und über KI: Chancen für eine Reformierung der Bildung, in: Achour, Sabine / Bieling, Hans / Massing, Peter / Schieren, Stefan / Varwick, Johannes (Hg.): *Künstliche Intelligenz*, POLITIKUM, Frankfurt/Main 2021, 54–61.
- Nordmann, Alfred: *Technikphilosophie. Zur Einführung*, Hamburg 2008.
- Otte, Roland: *Künstliche Intelligenz für dummies*, Weinheim 2019.
- Precht, Richard David: *Freiheit für alle. Das Ende der Arbeit wie wir sie kannten*, München 2022.
- Precht, Richard David: *Künstliche Intelligenz und der Sinn des Lebens*, München 2020.
- Renn, Jürgen: *Die Evolution des Wissens. Eine Neubestimmung der Wissenschaft für das Anthropozän*, Berlin 2022.
- Riehm, Tobias: Roboter als Personen im Rechtssinne? Zur Diskussion um eine digitale Rechtspersönlichkeit, in: *A VISO. Magazin für Kunst und Wissenschaft in Bayern*, 01/20, 2020, 30–33.
- Robinsohn, Saul Benjamin: *Bildungsreform als Revision des Curriculums*, 2. Aufl., Neuwied und Darmstadt 1979.
- Robinsohn, Saul Benjamin: Neue Lern- und Lehrinhalte, in: Brauneiser, Max (Hg.): *Attacken auf die Pädagogische Provinz*, Stuttgart 1970, 65–74.
- Ropohl, Günter: *Technologische Aufklärung. Beiträge zur Technikphilosophie*, 2. Aufl., Frankfurt am Main 1999.
- Schauer, Christian / Frank, Ulrich: Wirtschaftsinformatik an Schulen. Status und Desiderata mit Fokus auf Nordrhein-Westfalen, online verfügbar unter: https://www.icb.wiwi.uni-due.de/fileadmin/fileupload/ICB/research/research_reports/ICB-Report-No61.pdf, 2014 (abgerufen am 10.12.2024).
- Schwill, Andreas: *Fundamentale Ideen der Informatik*, 1993.
- Stalder, Felix: *Kultur der Digitalität*, Berlin 2016.
- Yogeshwar, Ranga: *Nächste Ausfahrt Zukunft. Geschichten aus einer Welt im Wandel*, Köln 2019.

Verzeichnis der AutorInnen und InterviewpartnerInnen

Prof. Dr. Daniel Cremers ist Ordinarius für Bildverarbeitung und Künstliche Intelligenz an der Technischen Universität München und dort unter anderem Direktor des Munich Center for Machine Learning.

Dr. Dr. Mariusz Chrostowski ist wiss. Assistent am Lehrstuhl für Didaktik der Religionslehre, für Katechetik und Religionspädagogik an der Theologischen Fakultät der Katholischen Universität Eichstätt-Ingolstadt.

Ornella Fieres ist eine auf neue Technologien spezialisierte Medienkünstlerin, deren Werke weltweit an renommierten Ausstellungsorten präsentiert werden. Sie hält international Vorträge und verfolgt eine Promotion im Bereich Medienwissenschaften an der Hochschule für Gestaltung Offenbach.

Prof. Dr. Johannes Heger ist Inhaber des Lehrstuhls für Religionspädagogik und Didaktik des Religionsunterrichts an der Katholisch-Theologischen Fakultät der Universität Julius-Maximilians-Universität Würzburg.

PD Dr. Michael Köck ist Fachvertreter für Arbeitslehre und Wirtschaftsdidaktik an der Katholischen Universität Eichstätt-Ingolstadt.

Dr. Christoph Kreisbeck ist Gründer und CEO von Aixelo Inc., einem KI-Startup, das datenbasierte Entscheidungsprozesse in die Forschungs- und Entwicklungsabteilungen der Chemie- und Werkstoffindustrie einbindet.

Prof. Dr. Lukas Ohly ist außerplanmäßiger Professor für Systematische Theologie und Religionsphilosophie an der Goethe-Universität Frankfurt am Main.

Jun.-Prof. Dr. Christina Pfeuffer ist Juniorprofessorin für Human-Technology Interaction an der Katholischen Universität Eichstätt-Ingolstadt.

Künstliche Intelligenz revolutioniert unseren Alltag. Dieser Band bietet eine fundierte Reflexion über die rasanten Entwicklungen von KI. Er vereint wissenschaftliche Aufsätze aus Theologie, Philosophie, Kunst und Didaktik mit spannenden Experteninterviews. Ein Buch für alle, die KI-Technologien nicht nur nutzen, sondern deren gesellschaftliche Folgen kritisch verstehen wollen.

Dr. Tobias Holischka ist Projektreferent am Zentrum für Forschungsförderung und Studiengangskordinator an der Fakultät für Soziale Arbeit der KU. Er war von 2009 bis 2022 wissenschaftlicher Mitarbeiter am Lehrstuhl für Philosophie der KU Eichstätt-Ingolstadt.

Prof. Dr. Carolin Kreisbeck ist Inhaberin des Lehrstuhls für Mathematik – Analysis an der KU Eichstätt-Ingolstadt.

Jun.-Prof. Dr. Christina Pfeuffer ist Inhaberin der Professur für Human-Technology Interaction an der KU Eichstätt-Ingolstadt.

Dr. Michael Winkmann ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Religionspädagogik an der KU Eichstätt-Ingolstadt.

ISBN 978-3-534-64351-6



www.herder.de