

KI-VISUALISIERUNGEN ALS IMPULS FÜR LITERARISCHES LERNEN

DIDAKTISCH-METHODISCHE ÜBERLEGUNGEN ZUM EINSATZ BILDGENERIERENDER KI AM BEISPIEL LYRI- SCHER TEXTE

Anastasia Keppler

Goethe-Universität Frankfurt am Main | a.keppler@em.uni-frankfurt.de

ABSTRACT

Der Beitrag widmet sich dem Einsatz bildgenerierender KI im Literaturunterricht mit dem Fokus auf dem literarischen Lernen. Ausgangspunkt ist die didaktische Tradition von Visualisierungen, die durch KI-gestützte Bildproduktion eine neue Dimension erhält: Lernende können Texte über Prompts unmittelbar in visuelle Darstellungen übersetzen und diese in den Reflexionsprozess einbeziehen. Nach einer theoretischen Einordnung von Visualisierungen und einer Beschreibung der Funktionsweise von Text-zu-Bild-Modellen werden aus deren technischen Bedingungen zentrale didaktisch-methodische Konsequenzen abgeleitet. Darauf aufbauend systematisiert der Beitrag ein Spektrum von Prompt-Typen, beschreibt die jeweils zu erwartenden Bildausgaben und skizziert exemplarisch anhand lyrischer Texte unterschiedliche Einsatzszenarien im Unterricht. Im Fazit wird schließlich die Übertragbarkeit auf epische Texte mit ihren spezifischen Anforderungen aufgezeigt.

SCHLAGWÖRTER

— BILDGENERIERENDE KI — VISUALISIERUNG — PROMPTING — LYRIKUNTERRICHT

ABSTRACT (ENGLISH)

**AI-GENERATED VISUALIZATIONS AS A STIMULUS FOR LITERARY
LEARNING. DIDACTIC CONSIDERATIONS ON THE USE OF TEXT-TO-IMAGE
GENERATION IN THE CONTEXT OF POETRY**

This article explores the use of AI-based image generation in literature education with a focus on literary learning. It starts from the didactic tradition of visualizations, which gains a new dimension through text-to-image models: learners can directly transform texts into visual representations via prompts and integrate them into processes of reflection. After situating visualizations theoretically and describing the functioning of text-to-image systems, the article derives key didactic-methodological implications from their technical conditions. Building on this, the paper systematizes a spectrum of prompt types, outlines the image outputs that can be expected, and illustrates exemplary classroom scenarios with poetry. In conclusion, it highlights the transferability of these approaches to narrative texts, taking into account their specific requirements.

KEYWORDS

— TEXT-TO-IMAGE AI — VISUALIZATION — PROMPTING — POETRY

1 — EINLEITUNG

Mit dem Einzug generativer KI-Tools in Bildungskontexte hat sich das Spektrum digitaler Werkzeuge im Deutschunterricht erweitert. Neben großen Sprachmodellen, die die Schreibprozesse verändern, finden auch bildgenerierende KI-Modelle Einzug in Klassenzimmer. Erste unterrichtspraktische Ansätze nutzen diese Technik vor allem zur Erstellung illustrativer Bilder für Arbeitsblätter (vgl. Deutsche Telekom Stiftung 2024), als kreative Schreibimpulse (vgl. Teachtoday 2023) oder im Kontext kritischer Medienbildung (vgl. Flick 2023), etwa durch die Analyse von KI-generierten Bildern im Hinblick auf Verzerrungen, Bias und algorithmische Bildlogiken. Parallel dazu wird ‚Prompting‘ als neue sprachlich-funktionale Routine profiliert (vgl. Junges 2024). In den bisherigen didaktischen Impulsen überwiegt damit der Fokus auf der Unterstützung von Schreibkompetenz und anwendungsbezogener Medienkompetenz, während die Potenziale der Technologie für literarische Lernprozesse bislang nur randständig mitgedacht werden.

Dabei ist die Arbeit mit Visualisierungen im Deutschunterricht keineswegs neu: Illustrationen in Schullesebüchern, Tafelbilder, Mindmaps oder auch künstlerische Bild-Text-Formate wie Bilderbücher, Comics und Graphic Novels sind seit langem Teil der Unterrichtspraxis. Bildgenerierende KI bringt eine zusätzliche Dimension in diese Praktiken: Bilder entstehen nicht mehr durch Auswahl oder handwerkliche Gestaltung, sondern durch sprachliche Steuerung im Prompt, und werden dadurch eng mit der Textarbeit der Lernenden verknüpft. Diese operative Kopplung von Sprache und Bild birgt besonderes didaktisches Potenzial: Lernende können ihre Deutungen in eine visuelle Form überführen, Unterschiede zwischen Text und Bild reflektieren und zugleich die Bedingungen medialer Produktion kritisch mitdenken.

Vor diesem Hintergrund rückt der Beitrag die Frage ins Zentrum, wie sich bildgenerierende KI gezielt für literarisches Lernen nutzen lässt. Um dieser Frage nachzukommen, werden zuerst Anknüpfungspunkte bei den bestehenden Praktiken der Bildproduktion im Deutschunterricht gesucht, KI-generierte Bilder werden damit in den Diskurs zur Arbeit mit Visualisierungen eingeordnet. Im Weiteren wird die Funktionsweise der Bildgeneratoren in Grundzügen dargestellt, um daraus didaktisch und methodisch relevante Konsequenzen für den Unterrichtsgebrauch abzuleiten. Darauf aufbauend differenziert der Beitrag konkrete Einsatzmöglichkeiten: Entlang eines Spektrums von Prompt-Typen werden verschiedene Szenarien literarischer Textarbeit skizziert, in denen KI-Visualisierungen eine produktive Wirkung entfalten können.

Die Beispiele und Szenarien konzentrieren sich exemplarisch auf Lyrik. Die hohe sprachliche Verdichtung und die ausgeprägte Bildlichkeit bieten ein fruchtbares Terrain, um die Potenziale und Grenzen der neuen Form der Visualisierungen zu erkunden. Die grundlegenden Verfahren und Beobachtungen sind auch auf epische und dramatische Texte übertragbar. Da diese mitunter andere spezifische Merkmale aufweisen, wird im Ausblick skizziert, unter welchen Bedingungen und mit welchen didaktischen Anpassungen ein solcher Transfer sinnvoll erfolgen kann.

2 — VISUALISIERUNGEN IM DEUTSCHUNTERRICHT

Im Deutschunterricht spielt die visuelle Dimension literarischer Texte seit jeher eine wichtige Rolle für die Textaneignung. Die rezeptive Auseinandersetzung mit bestehenden Bild-Text-Kombinationen wie Bilderbüchern und Graphic Novels ist dabei ebenso Teil der Unterrichtspraxis wie die produktive visuelle Gestaltung von eigenen Inhalten. Tafelbilder, Sketchnotes, Mindmaps und Collagen begleiten Lernprozesse, indem sie Informationen aufbereiten und strukturieren.

Die Praktiken, die den Ausgangstext mit einer grafischen Repräsentation verbinden, werden unter ‚Visualisierung‘ zusammengefasst. Der Begriff ist dabei weit gefasst und umfasst sowohl visuelle Produkte als auch den Prozess ihrer Erstellung und Reflexion. Gerhard Eikenbusch definiert Visualisierungen als

Aufbereiten, Konstruieren, Anwenden, Interpretieren und Reflektieren von piktorialen, grafischen oder visuellen Repräsentationen von Mitteilungen, Aussagen, Daten, Informationen und Zusammenhängen mit dem Ziel, Informationen zu erschließen und zu kommunizieren, Probleme zu lösen, Ideen zu entwickeln, mentale Modelle aufzubauen und vertieftes Verständnis zu erreichen (Eikenbusch 2024, 13).

In Anschluss an Norma Presmeg (2006) fassen Petra Gretsche und Lars Holzäpfel Visualisierungen als Repräsentationen, die von Lernenden interpretiert, verwendet und eigenständig konstruiert werden und „einen spezifischen didaktischen Zweck erfüllen“ (Gretsche / Holzäpfel 2016, 9). Tatjana Jesch und Michael Staiger unterscheiden zudem zwischen internen (intramentalen) sowie externen (extramentalen) Visualisierungen (vgl. Jesch / Staiger 2016, 63). Zusammengefasst markieren diese Bestimmungen Visualisierungen als didaktisch gerahmte Repräsentationen, die kognitive wie kommunikative Prozesse gezielt unterstützen.

Die Vorteile der Arbeit mit Visualisierungen für das literarische Lernen sind kognitionspsychologisch und didaktisch begründet. Nach der Dual-Coding-Theorie (vgl. Paivio 1986) und dem integrativen Modell des Text-Bild-Verstehens (vgl. Schnitz / Bannert 2003) nutzen Lernende bei der Auseinandersetzung mit Text-Bild-Kombinationen getrennte, aber kooperierende kognitive Systeme; die Kombination sprachlicher und visueller Information erleichtert Verstehen und Behalten und fördert die Ausbildung tragfähiger mentaler Modelle. Empirisch ist die positive Wirkung externer Visualisierungen für die Behaltens- und Verstehensleistung belegt (vgl. Guo et al. 2020; Sadoski / Paivio 2001); besonders wirksam für das Textverstehen erweist sich die Kombination interner und externer Visualisierungen (vgl. Gambrell / Jawitz 1993; Sadoski / Paivio 2001, 182, zit. nach Jesch / Staiger 2016, 71).

KI-generierte Bilder lassen sich als besondere Form externer Visualisierung begreifen: Sie entstehen in enger Kopplung an einen Text und werden in die Textarbeit eingebunden. Die Erzeugung verschiebt sich dabei von ‚handwerklichen‘ Verfahren hin zur sprachlichen Steuerung der Bildproduktion – im Sinne einer „operative[n] ekphrasis“ (Bajohr 2024, 80), bei der sprachliche Beschreibungen unter Rückgriff auf Wissen über Funktionsweisen der Modelle Bilder hervorbringen. Damit eröffnen KI-Visu-

alisierung einerseits Anschlussmöglichkeiten an etablierte Visualisierungspraxen, implizieren andererseits aber Kompetenzverschiebungen und neue Anforderungen an Lernende. Ein zentraler Bezugspunkt ist die handlungs- und produktionsorientierte Literaturdidaktik (vgl. Jesch / Staiger 2016, 67), in der interne wie externe Visualisierungen dazu dienen, Vorstellungsbilder zu konkretisieren, Deutungsprozesse anzustoßen und alternative Perspektiven auf literarische Texte zu eröffnen. In methodischer Hinsicht greifen KI-generierte Visualisierungen diese Intention auf. Durch Künstliche Intelligenz, die mit eigenen Daten, Filtern und Steuerungslogiken eine zusätzliche Instanz in den Prozess einführt, rückt diese Art der Visualisierung näher an rezeptive oder reflexive Verfahren, bei denen die Konfrontation mit einer fremden Visualisierung Ausgangspunkt für Diskussion, Irritation oder Interpretation ist.

3 — TEXT-ZU-BILD-KI: FUNKTIONSWEISE UND DIDAKTISCH-METHODISCHE IMPLIKATIONEN

Text-zu-Bild-KI-Modelle ermöglichen es, aus einer Eingabe in natürlicher Sprache (*Prompt*) eine visuelle Darstellung zu erzeugen. Sie bestehen heute meist aus einer Abfolge spezialisierter Komponenten. Zunächst wandelt ein Text-Encoder (z.B. CLIP) die Texteingabe in eine numerische Repräsentation (*Vektor*) im sogenannten Einbettungsraum um. In diesem Raum sind Texte und Bilder so codiert, dass semantisch Ähnliches nah beieinanderliegt. Die so erzeugte Text-Einbettung dient dem nachfolgenden Bildgenerator als Leitsignal für den Bildaufbau. Das eigentliche Bild entsteht in einem *Diffusionsmodell* (z.B. GLIDE), das die eingebettete Prompt-Repräsentation schrittweise in eine visuelle Struktur übersetzt. Dieser Prozess erfolgt iterativ, wobei das Modell in jedem Schritt neue Bildmerkmale generiert und sie an der semantischen Ausrichtung des Prompts orientiert. Häufig folgen weitere Module: Sogenannte *Verfeinerer* reichern das Bild mit Details an, *Upscaler* erhöhen die Auflösung, *Moderations filter* blockieren problematische Inhalte und sogenannte *Reranker* wählen ästhetisch oder inhaltlich besonders passende Bildvarianten aus (vgl. Linde 2023; Nöcker-Prior 2023).

Trainiert werden diese Systeme auf sehr großen Sammlungen von Bild-Text-Paaren. Diese Datensätze entstehen entweder durch gezielte Annotation (beispielsweise im Fall von COCO durch Crowdworke:innen, die Bildunterschriften verfassen (vgl. Lin et al. 2014)), oder sie werden automatisiert aus Webinhalten gewonnen (wie etwa im Fall des Datensatzes LAION-5B, der auf öffentlich zugänglichen Bild-Text-Kombinationen basiert (vgl. Beaumont 2022)). Auch Bewertungsmodule, etwa für die Auswahl ‚gelungener‘ Bilder, greifen auf Modelle zurück, die an menschlichen Präferenzen trainiert wurden.

Für die praktische Anwendung ist zudem zwischen eigenständigen Bildgeneratoren und deren Integration in Chatumgebungen zu differenzieren. Wird die Bildgenerierung über einen Chatbot angesteuert – etwa bei ChatGPT in Verbindung mit DALL·E 3 –, muss die Bildausgabe explizit aktiviert werden. In solchen Fällen agiert das Sprachmodell als vermittelnde Instanz: Es reformuliert den Prompt nach eigenem ‚Verständnis‘ für den angeschlossenen Bildgenerator, wodurch sich die Bildresultate je

nach Sprachmodell unterscheiden können. Da diese zwischengeschaltete Instanz die direkte Steuerbarkeit und Nachvollziehbarkeit des Promptings beeinflusst, konzentriert sich die folgende didaktische Auseinandersetzung auf die Nutzung externer Text-zu-Bild-Generatoren.

Aus den skizzierten technischen Grundlagen ergeben sich mehrere Hinweise für die unterrichtliche Nutzung. Erstens ist zu berücksichtigen, dass die Modelle Anweisungen nicht semantisch ‚verstehen‘, sondern lediglich statistische Korrelationen zwischen Sprache und Bildmustern reproduzieren. Entsprechend werden nicht alle sprachlichen Vorgaben gleich zuverlässig umgesetzt. Lernende sollten darauf vorbereitet werden, um Enttäuschungen über vermeintliches ‚Fehlverhalten‘ der Modelle zu vermeiden. Aus didaktischer Perspektive fördert die Notwendigkeit, Prompts präzise zu formulieren, eine eigenständige Kompetenz, die die sprachliche Ausdrucksweise der Lernenden schärft (vgl. Federiakin 2024).

Zweitens bestimmen die Trainingsdaten maßgeblich die Bildresultate. Systeme reagieren besonders stark auf Begriffe, die in den Daten häufig vorkommen und visuell leicht darstellbar sind (z.B. konkrete Substantive oder Stilmarker), und deutlich schwächer auf abstrakte Relationen, implizite Bedeutungen oder metaphorische Sprache. Da die Datensätze auf menschlich produzierten Beschreibungen und Bewertungen basieren, fließen dabei kulturelle und sprachliche Konventionen ebenso ein wie unbewusste Stereotype (vgl. Wan et al. 2024). Hier sind eine didaktische Begleitung und kritische Reflexion unerlässlich.

Drittens greifen viele Systeme auf Moderationsfilter zurück, die Texteingaben blockieren, die in literarischen Kontexten eigentlich unproblematisch wären. So können selbst gängige Motive des Barock wie *Memento mori* zur Ablehnung des Prompts führen, da Begriffe, die mit Tod in Verbindung stehen, als potenziell jugendgefährdend eingestuft werden. Da die genauen Filterkriterien nicht transparent sind, lässt sich nicht zuverlässig vorhersagen, auf welche Formulierungen ein System reagiert. Neben einem vorbereitenden Erproben geplanter Aufgaben kann dies didaktisch produktiv genutzt werden: Ablehnungen oder Blockierungen lassen sich als Anlass nehmen, Moderationslogiken zu thematisieren. Als Alternative kann zudem die Nutzung von Bildgeneratoren in Chatumgebungen sinnvoll sein, da dort das vorgeschaltete Sprachmodell Prompts teilweise umformuliert und dadurch andere Filterentscheidungen ausgelöst werden.

Viertens sind Nicht-Determinismus und Blackbox-Charakter inhärente Eigenschaften dieser Systeme. Identische Prompts können je nach Modell, Version oder Voreinstellung zu unterschiedlichen Bildausgaben führen, und dieselben Ergebnisse lassen sich oft nicht exakt reproduzieren. Für die schulische Praxis erfordert dies eine gewisse Flexibilität im Umgang mit den Resultaten und eine methodische Rahmung, die die Differenzen produktiv macht, statt sie als Defizit zu bewerten.

Darüber hinaus ist zu berücksichtigen, dass bildgenerierende KI-Systeme in komplexe Infrastruktur- und Machtverhältnisse eingebunden sind. Dies betrifft zum einen den erheblichen Ressourcenverbrauch datenintensiver Rechenzentren (vgl. Strubell / Ganesh / Callum 2019) und zum anderen die oft prekären Arbeitsbedingungen von

Crowdworker:innen, die weltweit für die Sortierung, Annotation und Bewertung von Daten eingesetzt werden (vgl. Crawford 2024). Für den Unterricht ergibt sich daraus die Notwendigkeit, Nutzen und Grenzen dieser Modelle nicht allein unter methodischen Gesichtspunkten zu betrachten. Vielmehr erscheint eine kritische Reflexion geboten, die auch ökologische, rechtliche und arbeitsethische Aspekte in den Blick nimmt.

4 — POTENZIALE FÜR DAS LITERARISCHE LERNEN: EINE PROMPT-TYPOLOGIE

Text-zu-Bild-KI eröffnen eine Reihe didaktisch anschlussfähiger Einsatzmöglichkeiten im Literaturunterricht und können das Repertoire etablierter methodischer Ansätze erweitern. Im Folgenden wird der Versuch unternommen, systematisch zu erfassen, welche Funktionen KI-Visualisierungen in unterschiedlichen Phasen der literarischen Textarbeit übernehmen können und welche Aspekte des literarischen Lernens dabei angesprochen werden.

Im Zentrum stehen Szenarien, in denen die Bildgenerierung durch die Lernenden selbst erfolgt. Zwar lässt sich auch eine Verwendung durch die Lehrkraft im Rahmen der Unterrichtsvorbereitung vielfältig gestalten, in der unterrichtlichen Nutzung unterscheidet sich ein solches Vorgehen jedoch nicht grundsätzlich von der Arbeit mit konventionellen Illustrationen oder künstlerischen Bildern. Das Spezifische des KI-Einsatzes liegt vielmehr in der direkten Verbindung der Bildgenerierung mit der rezeptiven und produktiven Auseinandersetzung der Lernenden mit dem Text.

Zur Strukturierung der Einsatzmöglichkeiten wird im Folgenden nach Typen der zur Bildgenerierung genutzten Prompts differenziert. Zentrales Kriterium ist dabei der Grad der Auseinandersetzung mit dem literarischen Text, der zur jeweiligen Promptformulierung vorausgesetzt wird. Daraus ergibt sich folgendes Spektrum: Am unteren Ende steht die direkte Nutzung des literarischen Textes oder eines Textausschnitts als Prompt, die noch keine Beschäftigung mit dem Text voraussetzt. Am oberen Ende finden sich interpretative Bildbeschreibungen, die auf einer detaillierten Analyse und Deutung des Textes beruhen und eine differenzierte bildliche Vorstellung erfordern. Dazwischen liegen Formen, die etwa mit ersten Eindrücken, symbolischen Elementen oder Schlüsselwörtern arbeiten und noch keine explizite Interpretation verlangen.

Im Folgenden werden drei exemplarische Prompt-Typen vorgestellt: Unbearbeiteter Text (ausschnitt), Schlüsselwörter zum Text und die interpretative Bildbeschreibung. Für jede Variante wird dargestellt, wie die daraus entstehenden Bilder typischerweise ausfallen, welche didaktischen Settings sich für den Einsatz anbieten und welche Kompetenzfelder durch diese Art des Einsatzes angesprochen werden.

4.1 UNBEARBEITETER TEXT (AUSSCHNITT)

Bei dieser Methode dient der literarische Text selbst – vollständig oder in Auszügen – als direkter Prompt für die KI-Bildgenerierung. Dieser Ansatz ist besonders

niedrigschwellig und eignet sich als Einstieg in die Arbeit mit KI-Visualisierungen. Aus technischen Gründen empfiehlt es sich bei längeren Gedichten, die Eingabe auf einzelne Strophen zu beschränken, um verwertbare Ergebnisse zu erhalten. Dieser Prompt-Typ erfordert keine Vorarbeit seitens der Lernenden; auf eine inhaltliche oder stilistische Steuerung wird hier bewusst verzichtet.

Die so erzeugten Visualisierungen orientieren sich typischerweise stark an der wörtlichen Ebene des Textes. Die KI-Modelle identifizieren meist zentrale Substantive und atmosphärische Begriffe und setzen diese bildlich um. Dabei besteht eine Tendenz, frühe Abschnitte des Prompts stärker zu gewichten, sodass spätere Motive mitunter unterrepräsentiert bleiben. Auch werden Objekte aus dem Text zwar übernommen, ihre (räumlichen) Beziehungen zueinander aber nicht immer korrekt dargestellt. Die stilistische Ausrichtung der Bilder wird zum großen Teil von den im Ausgangstext eingeschriebenen Hinweisen bestimmt, ist aber von Trainingsdaten des Modells beeinflusst. Bei identischem Prompt ist die Varianz der Ergebnisse innerhalb eines Modells in der Regel gering. Auffällig ist, dass viele Generatoren den Input als Gedicht erkennen und dies typografisch sichtbar machen – manchmal sogar ergänzt durch Autorenangaben oder Layouts, die an Buchillustrationen erinnern (Abb. 1)¹. In diesem Sinne übernehmen die erzeugten Bilder vor allem eine „veranschaulichende und wissensgenerierende Funktion“ (Gretsch / Holzäpfel 2016, 14): Sie machen sprachliche Oberflächenmerkmale sichtbar und eröffnen durch ihre Eigenlogik neue Deutungsanlässe.



Abb. 1: Bild zu Christian Morgensterns *Das große Lalulā*; vollständiger Gedichttext als Prompt

Didaktisch eignet sich dieser Typ vor allem für die initiale Auseinandersetzung mit einem literarischen Text. Obwohl der Ansatz keine tiefergehende Textanalyse voraussetzt, sollten die Lernenden vor der Bildgenerierung stets die Gelegenheit erhalten, eigene Vorstellungen zu entwickeln. Eine Visualisierung vor der ersten Lektüre kann die subjektive Vorstellungskraft verengen. Empirische Befunde weisen darauf hin, dass Lernende nach der Lektüre illustrierter Texte ihre inneren Bilder nicht von vorgegebenen Visualisierungen unterscheiden und nicht-illustrierte Texte tendenziell mehr Vorstellungsbildung evozieren (vgl. Sadoski / Paivio 2001, 108; zit. nach Jesch / Staiger 2016, 70). Erfolgt der Einsatz hingegen nach dem ersten Lesen, kann das KI-

¹ Abb. 1 wurde mit DALL·E 3 über ChatGPT erzeugt, alle weiteren Abbildungen wurden mit LeonardoAI über SchulKI generiert. Alle Abbildungen wurden im März 2026 erstellt.

Bild als Kontrastfolie zu den bereits entstandenen internen Visualisierungen dienen. In der Gegenüberstellung von eigener Vorstellung und externer Visualisierung werden semantische Schwerpunkte und Auslassungen deutlich. Anhand dieser Differenzen und Ähnlichkeiten lässt sich die Deutungsarbeit anbahnen und der Übergang zu einer vertiefenden Textarbeit schaffen, indem die Lernenden ihre Beobachtungen argumentativ an Textstellen zurückbinden (vgl. Keppler 2025).

Als produktiver Irritationsanlass kann die Tendenz der KI zur wörtlichen Konkretisierung von Metaphern dienen. So enthalten beispielsweise viele KI-Visualisierungen zu Eduard Mörikes Gedicht *Er ist's* (Mörike 1976, 684) ein reales „blaues Band“ (Abb. 2), während bei Goethes *Willkommen und Abschied* (Goethe 1998, 837) mitunter hunderte Augen die Betrachtenden anstarren (Abb. 3). Solche direkten Umsetzungen können die Aufmerksamkeit auf die entsprechenden sprachlichen Bilder lenken und eine tiefere Auseinandersetzung mit deren übertragener Bedeutung anstoßen.

Besonders fruchtbar kann dieser Prompt-Typ auch bei Gedichten sein, die kaum visuelle Anhaltspunkte bieten, wie etwa in der Lautpoesie. Bei solchen Ausgangstexten können die Ergebnisse sehr unerwartet ausfallen, auch die Varianz der Visualisierungen erhöht sich signifikant. Hier können die generierten Bilder neue Deutungshorizonte eröffnen und als Impuls für die Interpretation dienen. Während



Abb. 2: Bild zu Eduard Mörikes *Er ist's*; vollständiger Gedichttext als Prompt



Abb. 3: Bild zu Johann Wolfgang von Goethes *Willkommen und Abschied*; erste Strophe als Prompt

Christian Morgensterns *Das große Lalulā* (Morgenstern 1975, 249) etwa in eine abstrakte, märchenhafte Ästhetik (Abb. 4) oder in eine Szene mit magischen Elementen (Abb. 5) überführt wird, resultiert Hugo Balls *Karawane* (Ball 1963, 28) entweder in grafischen Schrift-Kompositionen (Abb. 6) oder in Darstellungen, die die Situation einer Bühnenperformance abbilden (Abb. 7). Die Konfrontation mit solchen unterschiedlichen Ergebnissen kann die Reflexion über Text-Bild-Beziehungen anstoßen und die Pluralität möglicher Lesarten verdeutlichen.

Über die Textanalyse hinaus kann dieser Ansatz auch für die Förderung der KI-Kompetenz genutzt werden. Die Ergebnisse bieten eine gute Grundlage, um modellbedingte Effekte wie Stereotypisierung, den Fokus auf Oberflächenmerkmale oder die geringe Varianz zu diskutieren. Die Lernenden können reflektieren, was das Bild hervorhebt, was es vernachlässigt und welche technischen Gründe dies haben könnte.

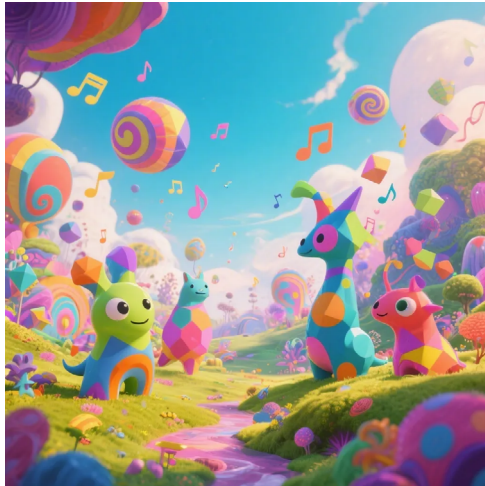


Abb. 4: Bild zu Christian Morgensterns *Das große Lalulä*; vollständiger Gedichttext als Prompt



Abb. 5: Bild zu Christian Morgensterns *Das große Lalulä*; vollständiger Gedichttext als Prompt



Abb. 6: Bild zu Hugo Balls *Karawane*; vollständiger Gedichttext als Prompt



Abb. 7: Bild zu Hugo Balls *Karawane*; vollständiger Gedichttext als Prompt

4.2 SCHLÜSSELWÖRTER

In dieser Variante identifizieren die Lernenden zentrale Begriffe, Motive oder Symbole des Gedichts und verwenden diese in kondensierter Form als Prompt. Voraussetzung ist mindestens eine erste Lektüre mit basaler Erschließung der themati-

schen Schwerpunkte; der Prompt setzt sich typischerweise aus Substantiven mit ausgewählten Attributen (gegebenenfalls Verben) zusammen. Abhängig von der Aufgabenstellung kann das Spektrum von eher abstrakten, stimmungsbezogenen Eindrücken bis zu konkreten, bildhaft anschlussfähigen Elementen reichen. Schon die Reduktion auf wenige Termini zwingt zu einer ersten Deutungsleistung: Lernende müssen den Text reduzieren und Begriffe gewichten, was begründet und am Wortlaut plausibilisiert werden muss; damit werden Auswahl- und Argumentationsfähigkeit früh im Prozess aktiviert.

Die resultierenden Bilder zeigen eine etwas größere Spannweite als bei der Eingabe des gesamten Textes, weil die Modelle mit weniger Vorgaben arbeiten und Leerstellen eigenständig schließen. Die stilistische Ausrichtung der Bilder hängt maßgeblich von den Voreinstellungen und Trainingsdaten des jeweiligen Modells ab. Dabei lassen sich zwei unterschiedliche Tendenzen identifizieren, die jeweils spezifische didaktische Reflexionsmöglichkeiten eröffnen: Einerseits führen stark bildlich-konkrete Prompts oft zu wenig überraschenden, objektzentrierten Ausgaben, da die Begriffe direkt in bildliche Repräsentationen überführt werden können. So führt ein konkret-objektzentrierter Schlüsselwort-Prompt wie ‚Mann, Frau, Café, Stock, Hut, Fenster, Schiff‘ bei Kästners *Sachlicher Romanze* (Kästner 1969, 111) zu einer additiven Anordnung der im Text genannten Realien (Abb. 8). Didaktisch lässt sich so das Verhältnis von wörtlicher und übertragener Sprache reflektieren: Die Visualisierung der Realien dient als Kontrastfolie zur intendierten Atmosphäre und zeigt, dass literarische Bedeutung über die bloße Objektdarstellung hinausgeht. Andererseits provozieren abstrakte Begriffe konventionalisierte Symbolik und kulturelle Standardmotive. So erzeugt der Rückgriff auf die abstrakteren Begriffe ‚Liebe, Trennung, Traurigkeit‘ eine klischeehafte Situation in herbstlicher Parkatmosphäre mit einer einsamen Figur (Abb. 9). Solche Standard-Darstellungen können durch gezielte didaktische Arrangements produktiv gewendet werden, indem sie als Anlass zur kritischen Reflexion über gewohnte Bildmuster und kulturelle Codes genutzt werden.



Abb. 8: Bild zu Erich Kästners *Sachliche Romanze*; Prompt: ‚Mann, Frau, Café, Stock, Hut, Fenster, Schiff‘



Abb. 9: Bild zu Erich Kästners *Sachliche Romanze*; Prompt: ‚Liebe, Trennung, Traurigkeit‘

Für den Unterricht bieten sich Arrangements an, die Auswahl- und Reduktionsleistungen explizit machen und die entstandene Varianz gezielt nutzen. Eine Möglichkeit besteht darin, mehrere Bilder zum identischen Schlüsselwort-Prompt zu vergleichen: Welche gemeinsamen Muster treten auf, welche Details unterscheiden sich, und welches Bild passt am ehesten zum Gedicht – und warum? In der Gegenüberstellung verschiedener Ausgaben zum Schlüsselwort-Prompt, 'Liebe, Trennung, Traurigkeit' (Abb. 9-11) fällt auf, dass in allen Bildern ein Paar abgebildet ist, obwohl nicht explizit im Prompt genannt; ähnlich ist auch der Rückgriff auf Natursymbolik in Form von welken Blättern oder Blumen. Unterschiede zeigen sich in der räumlichen Anordnung und Nutzung kultureller Codes: Während Abbildung 9 die Trennung über eine herbstliche Parkszene mit einer einsamen Figur im Vordergrund und einem sich entfernenden Paar im Hintergrund darstellt, flüchtet sich Abbildung 10 in eine klischeehafte Symbolik mit einer verwelkten Rose und zerrissenen Herzmotiven. Abbildung 11 wiederum nutzt Unschärfe und fliegende Vögel als ästhetische Zeichen für den Verlust, wobei das Paar hier noch in physischer Nähe verbleibt. Bei solchen Vergleichen müssen die Lernenden die visuellen Ergebnisse präzise beschreiben und ihre Beobachtungen argumentativ an Textstellen zurückbinden. Zugleich wird die Reflexion von Text-Bild-Bezügen geschärft, indem die Passung zwischen den ausgewählten Begriffen, der Bildausgabe und der Textbedeutung geprüft wird.



Abb. 10: Bild zu Erich Kästners *Sachliche Romanze*; Prompt: ‚Liebe, Trennung, Traurigkeit‘

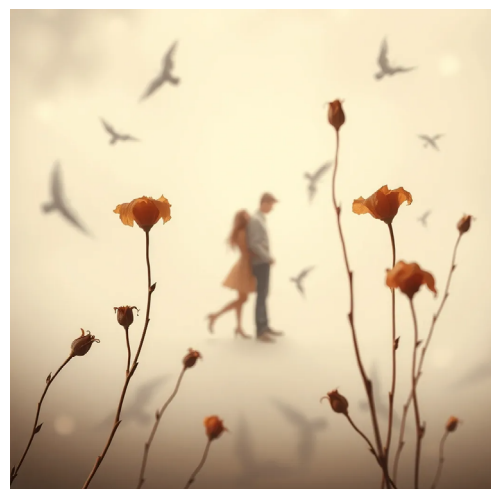


Abb. 11: Bild zu Erich Kästners *Sachliche Romanze*; Prompt: ‚Liebe, Trennung, Traurigkeit‘

Eine weitere Möglichkeit ist ein iteratives Vorgehen. Ausgehend von der Frage „Was fehlt?“ kann der Prompt gezielt erweitert oder umgeschrieben werden, um die Visualisierung an die eigene Vorstellung anzunähern. Ergänzen die Lernenden zum Beispiel den abstrakten Prompt zu Kästners *Sachlicher Romanze* etwa um den Ort („Café“) und den atmosphärischen Kern („Schweigen“), verschiebt sich die Darstellung zwar in den Innenraum, das Modell füllt die Leerstelle der Akteur:innen jedoch möglicherweise noch eigenständig und textfern – etwa durch eine einsame ältere Figur (Abb. 12). Erst die gezielte Hinzufügung des Begriffs ‚Paar‘ führt zu einer Visualisierung, die die räumliche und emotionale Distanz zwischen den Protagonisten im Café-Kon-

text einfängt (Abb. 13). Durch solche iterative Arbeit werden implizite Vorstellungen sprachlich präzisiert. Zugleich eröffnet dieser Ansatz einen transparenten Blick auf die modellbedingten Einflüsse: Welche Trainingsdaten und Voreinstellungen stützen die Standardmotive, welche Darstellungen sind erwartbar, und wo bleiben die Bildräume trotz Promptvariation stabil? So wird neben der vertieften Arbeit am Gedicht auch ein methodisches Verständnis für die KI-Generierung gefördert.



Abb. 12: Bild zu Erich Kästners *Sachliche Romanze*;
Prompt: ‚Liebe, Trennung, Café, Schweigen‘



Abb. 13: Bild zu Erich Kästners *Sachliche Romanze*;
Prompt: ‚Liebe, Trennung, Café, Schweigen, Paar‘

4.3 INTERPRETATIVE BIBLBESCHREIBUNG

Bei dieser Variante formulieren die Lernenden auf der Grundlage ihrer eigenen Textdeutung eine Bildbeschreibung, die anschließend als Prompt verwendet wird (vgl. Schulze / Seewald / Werner 2025). Anders als bei den vorherigen Typen handelt es sich nicht um eine direkte Auswahl oder Reduktion von Textelementen, sondern um eine bewusste Übersetzung der eigenen Lesart in eine visuelle Sprache. Dieser Prompt-Typ kann dabei von einem einzelnen Satz bis zu einem kurzen Absatz reichen und neben inhaltlichen auch stilistische oder kompositorische Vorgaben enthalten. In diesem Modus werden nicht nur atmosphärische Eindrücke sichtbar gemacht, sondern explizite Deutungshypothesen, die im Bild eine anschauliche Gestalt annehmen. Die Beziehung zwischen Text und Bild ist dabei nicht bloß illustrativ, sondern spiegelt einen Reflexionsprozess über den Text. So kann eine Interpretation zu Rilkes *Der Panther* (Rilke 1996, 505) etwa den emotionalen Zustand des Tieres ins Zentrum rücken und dabei die Bildlichkeit des Gedichts weitgehend beibehalten (Abb. 14), während eine andere Lesart sich stärker von den Gedichtbildern entfernt und das Motiv der Lähmung auf eine menschliche Situation überträgt, um psychologische Isolation in einem modernen Kontext sichtbar zu machen (Abb. 15). Dieser Prompt-Typ birgt ein hohes kreatives Potenzial, zugleich aber auch das Risiko, dass die Distanz zum Text größer wird und die Bilder sich in freien Assoziationen von der Textgrundlage lösen, weswegen eine bewusste Rückkopplung an den Text stets notwendig ist.

Die erzeugten Bilder hängen in sehr starkem Maße von der Ausgestaltung des Prompts ab. Dabei ist zu berücksichtigen, dass gute Bildbeschreibungen nicht automatisch gute Prompts sind. Hier kommen die „Prompting-Skills“ (Junges 2024) der Lernenden als entscheidender Einflussfaktor auf die erzeugten Bilder dazu, die benötigt werden, um die Modelle gezielt zu steuern und die gewünschten Resultate zu erzielen. Im Unterschied zu den anderen Varianten können und sollen hier auch Stil, Komposition oder Farbgebung bewusst gesteuert werden, sodass die Voreinstellungen der Modelle in den Hintergrund treten. Die Visualisierungen fallen bei identischen Prompts innerhalb des gleichen Modells sehr ähnlich aus, sodass die Lernenden den Zusammenhang zwischen Formulierungsvarianten und Bildresultat unmittelbar erfahren.



Abb. 14: Bild zu Rainer Maria Rilkes *Der Panther*, Prompt: ‚Nahaufnahme eines Panthers hinter dicken, unscharfen Gitterstäben. Der Blick ist leer und müde. Entsättigte Farben, starke Schlagschatten.‘



Abb. 15: Bild zu Rainer Maria Rilkes *Der Panther*, Prompt: ‚Silhouette einer Person in einem kargen Raum vor einem Fenster. Die Schatten der Jalousien fallen wie breite Gitterstäbe über den Körper. Filmischer Stil, kühles blaues Licht.‘

Didaktisch eignet sich diese Form für vertiefende Phasen der Textarbeit, in denen die Entwicklung und Präzisierung eigener Interpretationen im Vordergrund steht. Ein iteratives Vorgehen kann dabei hilfreich sein: Die Lernenden passen ihren Prompt so lange an, bis das Bild ihrer eigenen Vorstellung entspricht. Damit reiht sich diese Arbeitsweise in die Tradition handlungs- und produktionsorientierter Methoden ein, die eigenständige Gestaltung als Medium der Textaneignung nutzen. Gleichzeitig erfordert diese Methode auch eine Begleitung, da die Lernenden ein Grundverständnis für die Funktionsweise von Bildgeneratoren benötigen, um Frustration vorzubeugen, da die Modelle auch sehr genaue Anleitungen nicht immer eins zu eins umsetzen. Die entstandenen Bilder können vor allem als Anlass für Anschlusskommunikation genutzt werden. Hier übernehmen die KI-Visualisierungen eine „kommunizierende Funktion zur Aufbereitung eines Sachverhalts für einen Anderen“ (Gretsch / Holzäpfel 2016, 14). Auf diese Weise werden die individuellen Interpretationen anschaulich gemacht und der Vergleich verschiedener Lesarten innerhalb der Lerngruppe erleich-

tert. Es kann fruchtbare Diskussionen über Textverständnis, Perspektivenvielfalt und subjektive Zugänge begünstigen, bei denen die Reflexion von Text-Bild-Bezügen, die Fähigkeit zur argumentativen Begründung und die ästhetische Wahrnehmung gleichermaßen gefördert werden.

5 — FAZIT UND AUSBLICK

Der Einsatz von Text-zu-Bild-KI bietet Lernenden neue Möglichkeiten, eigene Vorstellungen zu literarischen Texten in visuelle Repräsentationen zu überführen und so eine kritische Auseinandersetzung mit diesen Texten zu fördern. Dabei werden nicht nur kreative und gestalterische Kompetenzen angesprochen, sondern auch die Fähigkeit zur Bedeutungsreduktion, die Reflexion von Text-Bild-Bezügen und die präzise sprachliche Gestaltung. Die didaktische Besonderheit der Arbeit mit KI-Visualisierungen liegt in der produktiven Verknüpfung rezeptiver und produktiver Verfahrensweisen. Lernende rezipieren einen Text, erzeugen auf dieser Basis ein Bild über Sprache und reflektieren im Anschluss das Verhältnis von Text, Bild und ihrer eigenen Interpretation. Je nach Prompt-Typ verlagern sich die Anforderungen und Potenziale. Das Spektrum reicht von der Auseinandersetzung mit stereotypisierenden Bildangeboten bis hin zur eigenständigen, sprachlich gesteuerten Visualisierung individueller Deutungen.

Obwohl sich die im Beitrag entwickelten Beispiele auf lyrische Texte konzentrieren, lassen sich die vorgestellten Verfahren grundsätzlich auch auf narrative Texte übertragen. Für die Gestaltung entsprechender Szenarien sind jedoch einige Besonderheiten zu berücksichtigen. Während bei lyrischen Texten häufig ein Gesamtverständnis als Grundlage für die Visualisierung ausreicht, bietet sich bei epischen Texten eher eine fokussierte Arbeit mit einzelnen Szenen, Figuren oder Perspektiven an. Dafür sprechen sowohl pragmatische als auch strukturelle Gründe: Zum einen ist die Texteingabe in gängigen KI-Generatoren in der Länge begrenzt, sodass sich epische Texte in der Regel nicht vollständig eingeben lassen. Zum anderen lässt sich die zeitliche Abfolge, die für epische Texte charakteristisch ist, nur schwer in einem einzigen Bild abbilden. Daraus ergeben sich spezifische Anforderungen an die Aufgabenstellung: Statt Einzelbildern eignen sich hier eher Bildfolgen – etwa in Form von szenischen Standbildern oder Storyboards. Didaktisch eröffnet dies neue Möglichkeiten, etwa zur Erarbeitung von Handlungsverläufen, zur Erkundung von Figurenkonstellationen oder zur Darstellung unterschiedlicher Fokalisierungen. Technisch problematisch bleibt jedoch die fehlende Bildkohärenz: Text-zu-Bild-Modelle erzeugen bei jeder Anfrage neue Bilder, ohne Garantie auf Figuren- oder Motivkonsistenz. Dennoch können KI-Visualisierungen auch im Umgang mit epischen Texten als produktiver Katalysator für Deutungsprozesse wirken – insbesondere dann, wenn sie gezielt zur Irritation, Perspektivierung oder Hypothesenbildung eingesetzt werden.

QUELLENVERZEICHNIS

PRIMÄRQUELLEN

—— **Ball, Hugo (1963)**: *Gesammelte Gedichte*. Zürich: Arche. ——— **Goethe, Johann Wolfgang von (1998)**: *Sämtliche Werke. Briefe, Tagebücher und Gespräche*. Band 1: *Gedichte 1756-1799*. 2. Aufl. Frankfurt a.M.: DKV. ——— **Kästner, Erich (1969)**: *Gesammelte Schriften für Erwachsene*. Band 1: *Gedichte*. München u.a.: Droemer Knauer. ——— **Morgenstern, Christian (1975)**: *Ausgewählte Werke*. Leipzig: Insel. ——— **Mörike, Eduard (1976)**: *Sämtliche Werke*. Band 1. München: Winkler. ——— **Rilke, Rainer Maria (1996)**: *Sämtliche Werke*. Band 1: *Gedichte*. 4. Aufl. Frankfurt a.M.: Insel.

SEKUNDÄRQUELLEN

—— **Bajohr, Hannes (2024)**: *Operative ekphrasis. The collapse of the text/image distinction in multimodal AI*. In: *Word & Image*, H. 2 (2024), 77-90. <https://doi.org/10.1080/02666286.2024.2330335>. ——— **Beaumont, Romain (2022)**: *LAION-5B. A new era of open large-scale multi-modal datasets*. <https://laion.ai/blog/laion-5b/> [10.09.2026]. ——— **Crawford, Kate (2024)**: *Atlas der KI. Die materielle Wahrheit hinter den neuen Datenimperien*. München: C.H. Beck. ——— **Deutsche Telekom Stiftung (2024)**: *Künstliche Intelligenz in der Schule*. <https://www.telekom-stiftung.de/sites/default/files/files/Leitfaden-kompakt-KI-Schule.pdf> [10.03.2026]. ——— **Eikenbusch, Gerhard (2024)**: *Visualisieren im Deutschunterricht. Eine Einführung in Methoden und Praxisideen*. Hannover: Klett | Kallmeyer. ——— **Federiakina, Denis et al. (2024)**: *Prompt engineering as a new 21st century skill*. In: *Frontiers in Education* 9. <https://doi.org/10.3389/feduc.2024.1366434>. ——— **Flick, Manuel (2023)**: *KI-Bilder im Unterricht. Fake News erkennen und entlarven*. <https://www.manueelflick.de/blog/ki-bilder-und-fake-news> [10.03.2026]. ——— **Gambrell, Linda B. / Jawitz, Paula Brooks (1993)**: *Mental Imagery, Text Illustrations, and Children's Story Comprehension and Recall*. In: *Reading Research Quarterly*, H. 3 (1993), 264. Auch online: <https://www.jstor.org/stable/747998?origin=crossref> [29.06.2025]. ——— **Gretsch, Petra / Holzäpfel, Lars (2016)**: *Einleitung*. In: Dies. (Hg.): *Lernen mit Visualisierungen. Erkenntnisse aus der Forschung und deren Implikationen für die Fachdidaktik*. Münster: Waxmann, 9-18. ——— **Guo, Daibao et al. (2020)**: *Do You Get the Picture? A Meta-Analysis of the Effect of Graphics on Reading Comprehension*. In: *AERA Open*, H. 1. <https://doi.org/10.1177/2332858420901696>. ——— **Jesch, Tatjana / Staiger, Michael (2016)**: *Bilder und Visualisierungen in der Lese- und Literaturdidaktik*. In: Petra Gretsch / Lars Holzäpfel (Hg.): *Lernen mit Visualisierungen. Erkenntnisse aus der Forschung und deren Implikationen für die Fachdidaktik*. Münster: Waxmann, 63-82. ——— **Junges, Stefanie (2024)**: *Prompt-Engineering. Prozessmodell*. <https://fachportal.lernnetz.de/index.php/sh/faecher/fuer-alle-faecher/lernen-mit-ki/praesizes-prompten.html> [10.03.2026]. ——— **Kepler, Anastasia (2025)**: *Was KI in Gedichten sieht. Strategischer Einsatz bildgenerierender KI als Unterstützung für das literarische Verstehen*. In: Florian Hesse / Gerrit Helm (Hg.): *Künstliche Intelligenz im Deutschunterricht jenseits der Textproduktion (Themenheft)*. Universitätsbibliothek der Ruhr-Universität Bochum. <https://doi.org/10.46586/SLLD.442.533>. ——— **Lin, Tsung-Yi et al. (2014)**: *Microsoft COCO. Common Objects in Context*. In: David Fleet et al. (Hg.): *Computer vision - ECCV 2014. 13th European conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*. Cham: Springer, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48. ——— **Linde, Helmut (2023)**: *Künstliche Intelligenz. So funktionieren KI-Bildgeneratoren*. <https://www.golem.de/news/kuenstliche-intelligenz-so-funktionieren-ki-bildgeneratoren-2305-174436.html> [10.03.2026]. ——— **Nöcker-Prior, Philipp (2023)**: *Bildgenerierung mit Künstlicher Intelligenz*. https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-2223/Bildgenerierung_mit_KI.pdf [10.03.2026]. ——— **Paivio, Allan (1986)**: *Mental representations: a dual coding approach*. New York u.a.: Oxford Univ. Press. ——— **Presmeg, Norma (2006)**: *Research on visualization in learning and teaching mathematics*. In: Gutiérrez, Ángel / Boero, Paulo (Hg.): *Handbook of Research on the Psychology of Mathematics Education. Past, Present and Future*. Münster: Brill, 205-235. ——— **Sadoski, Mark / Paivio, Allan (2001)**: *Imagery and text. A dual coding theory of reading and writing*. Mahwah: Erlbaum Associates. ——— **Schnotz, Wolfgang / Bannert, Maria (2003)**: *Construction and interference in learning from multiple representation*. In: *Learning and Instruction*, H. 2 (2003), 141-156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8). ——— **Schulze, Tina / Seewald, Charlotte / Werner, Sybille (2025)**: *Ästhetische Zugänge zu Literatur durch KI am Beispiel des Jugendromans Wolf von Saša Stanišić*. In: Florian Hesse / Gerrit Helm (Hg.): *Künstliche Intelligenz im Deutschunterricht jenseits der Textproduktion (Themenheft)*. Universitätsbibliothek der Ruhr-Universität Bochum. <https://doi.org/10.46586/SLLD.442.531>. ——— **Strubell, Emma / Ganesh, Ananya / McCallum, Andrew (2019)**: *Energy and Policy Considerations for Deep Learning in NLP*. In: Korhonen, Anna / Traum, David / Lluís, Màrquez (Hg.): *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 3645-3650. <https://doi.org/10.18653/v1/P19-1355>. ——— **Teachtoday (2023)**: *Unterrichtsidee zu bildgenerierender KI*. In: *Projektidee. Kreativ mit KI*. https://www.teachtoday.de/Themen/Medienkompetenz/ChatGPT_Co/3318_Projektidee_Kreativ_mit_KI.htm [10.03.2026]. ——— **Wan, Yixin et al. (2024)**: *Survey of Bias In Text-to-Image Generation: Definition, Evaluation, and Mitigation*. <https://doi.org/10.48550/arXiv.2404.01030>.

ÜBER DIE AUTORIN

[Anastasia Keppler](#) ist wissenschaftliche Mitarbeiterin am Institut für die deutsche Literatur und ihre Didaktik an der Goethe Universität Frankfurt. Ihre Forschungsinteressen liegen im Bereich der digitalen Literatur und Lesens in der Kultur der Digitalität.