

Alisa Geiß

Aus Text wird Bild

Diskurse der Illustration zu Bildgeneratoren und Prompts

Abstract: Over the last two years, the third wave of artificial intelligence (AI) has emerged powerful tools for both artistic expression and scientific research. In design, image generators display an equivalent disruption to text generators, while the medium of text creates the new scope of writing prompts. This contribution discusses the ambivalences between text and image generators via two main theses: first about the potential of prompting and generated images as a medium of discourse; second, it examines the reasoning behind their tendency to be mystified. Prompts can be automated through text generators and used to generate images on a wider scale, making it a praxis in which precise wording becomes crucial for generating anything. Generated images practice a form of mimesis. With Generative Adversarial Networks (GANs) producing non-referential images, the discourse of illustration is concerning the obsolescence of authorship. AI is observed to manifest a dual character, serving as both tool and author, even though AI does not own the classification of a person and is not an author in legal terms. The utilization of text and image generators is intertwined with a mythicized notion due to their non-verifiable mode of operation. The myth of AI is a product of its indeterminacy.

In der Geschichte der Technik ist die Beziehung zwischen Mensch und Maschine von einer Vorstellung geprägt: Maschinen über die menschliche Sprache bedienen zu können. Von den Anfängen der Chatbots wie Joseph Weizenbaums ELIZA¹ über Cleverbot bis hin zu KI-Textgeneratoren wie ChatGPT ist dieses Konzept längst Realität. Durch eine Textebene können Aufgaben nicht nur mühelos kommuniziert werden, sondern ebenso wird durch Text in seiner Funktion als In- und Output eine Anwendung des Textgenerators als Metainstanz ermöglicht. Eine Entwicklung, die an die Turing-Maschine erinnert, welche die Theorie des Programmes auf seine multiple Anwendbarkeit erweiterte.²

Betrachtet man die frühen Ansätze der künstlichen Intelligenz als interdisziplinäre Vision, wollte sie das überwinden, was C.P. Snow 1959 provokant als „Zwei

1 Vgl. Joseph Weizenbaum, *Die Macht der Computer und die Ohnmacht der Vernunft*, Berlin 1978, 305–306.

2 Alan Turing, „On computable numbers, with an application to the Entscheidungsproblem“, *Proceedings of the London Mathematical Society*, Bd. 42(1), 1937, 230–265.

Kulturen“-Phänomen beschrieb:³ Eine Trennung der wissenschaftlichen Disziplinen in Geistes- und Naturwissenschaften. KI verbinde technische und humane Arbeitsweisen – ein Narrativ, welches bereits vor 70 Jahren unter der Kybernetik existierte.⁴ Ist mit der dritten Welle künstlicher Intelligenz von einer Disruption in beiden Arbeitswelten zu sprechen, so trägt ein schneller Wandel die Konsequenz einer Mythenbildung. Unklar ist der Umgang mit KI-Generatoren, unklar der gesellschaftliche Konsens zu Restriktionen. Versuche, über die KI-Lernprozesse aufzuklären, scheitern dabei an zwei Faktoren, an der privatisierten Software im Zuge von marktführenden Unternehmen sowie an einer unmöglichen Nachvollziehbarkeit der Algorithmen. Ein ganzheitliches Verständnis bleibt unerreichbar, wenn kein Einblick in die Funktionsweisen möglich ist. Das Arbeiten mit KI grenzt somit an eine immanente Unbegreiflichkeit.

Die Anwendung von Prompts, einer textlichen Aufgabenstellung für die KI, beschränkt sich dabei nicht nur auf die Textproduktion. Im Bereich der Illustration bringen Bildgeneratoren wie Midjourney oder StableDiffusion eine ebenbürtige Disruption. Textgeneratoren können hier für eine automatisierte Produktion von Prompts eingesetzt werden. Die KI-unterstützte Produktion von Bildern über Prompts sowie deren mögliche Kommerzialisierung werfen eine Reihe von Fragen auf. Wie verändert der Einsatz von KI die Bildrezeption? Besitzen Text- und Bildgeneratoren eine Autorenfunktion? Können Prompts als Schreibpraxis betrachtet werden?

Bereits diese wenigen Leitfragen verdeutlichen, dass Text und Bild im Kontext von KI erneut betrachtet werden müssen. Dafür ist es hilfreich, ein Verständnis der Funktionsweisen von Bildgeneratoren zu besitzen. Ziel dieses Beitrages ist es, die Ambivalenzen zwischen Text- und Bildgeneratoren über zwei Hauptthesen zu erläutern, zunächst über das Potential von generierten Bildern als Diskursmedium, dann über ihren immanenten Mythos.

³ Vgl. C.P. Snow, *The Two Cultures*, Cambridge 2001 [1959].

⁴ Exemplarisch zu nennen sind hier Norbert Wiener, *Cybernetics or Control and Communication of the Animal and the Machine*, Cambridge 1961, 133–155; Karl Steinbuch, *Automat und Mensch. Kybernetische Tatsachen und Hypothesen*, Berlin/Heidelberg 1965 oder Georg Klaus, „Kybernetik – eine neue Universalwissenschaft der Gesellschaft?“, in *Zur Kritik der bürgerlichen Ideologie*, hg. von Manfred Buhr, Berlin 1973.

1 Grundlagen eines Generative Adversarial Networks (GANs)

Das Erkennen und Klassifizieren von Zusammenhängen ist ein wesentlicher Bestandteil des Lernens. KI, die mit einer Texteingabe Bilder produziert, ist auf ein multimediales Identifizieren von Signifikat und Signifikant trainiert. Das Standardmodell eines Bildgenerators basiert auf einem Generative Adversarial Network (GAN), welches eine spezielle Methodik zur Klassifizierung anwendet. KI beschreibt hier das Trainieren eines künstlichen neuronalen Netzes auf das Lösen komplexer Aufgaben.

Neuronale Netze funktionieren als Informationsträger, die den gewünschten Input verarbeiten. Ein Neuron ist mit Neuronen aus einer anderen Ebene verbunden. In einem Perzeptron, einem vereinfachten neuronalen Netz, gibt es drei Ebenen: Die Input-, Hidden- und Output-Ebene.⁵ Jede Ebene von Neuronen besitzt eine unterschiedliche Aufgabe.

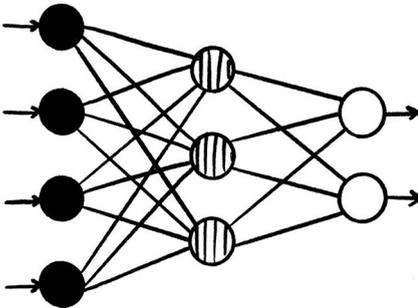


Abbildung 1: Visualisierung eines Perzeptrons mit Input-, Hidden- und Output-Ebene (v.l.n.r.).

Die Input-Ebene empfängt Informationen in einem beliebigen Dateiformat. Die Hidden-Ebene ist für die Formatierung und Verarbeitung der Information verantwortlich und bildet den Großteil des Netzes. Die Output-Ebene gibt die finalen Werte der Hidden-Ebene aus. Mehrere Ebenen von gleichartigen Neuronen bilden ein Layer.

Am Beispiel eines Kükens soll ein Netzwerk dieses auf einem beliebigen Bild erkennen. Wie gelingt das? Bei der Eingabe eines Bildes muss der Input-Layer das Bild erst dekodieren, um dessen Inhalt zu erfassen. Jedes Neuron übersetzt einen Pixel in einen numerischen Wert und sendet diesen Wert an den Hidden-Layer. Dort

⁵ Vgl. Günther Daniel Rey/Karl F. Wender; *Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*, Bern 2011, 13–14.

findet das Generieren von Zusammenhängen statt: Die Werte der Pixel werden so gewichtet, dass sie ein Muster bei gleichen Objekten aufzeigen. Befinden sich Bildelemente wie Schnabel, Augen, Körper oder Farbe des Kükens immer wieder im gleichen Verhältnis zueinander, werden sie höher gewichtet.⁶ Sind Bildelemente für die Identifikation irrelevant, wie Hintergründe oder andere Objekte, werden sie weniger gewichtet. Je stärker ein Netz gewichtet ist, desto präziser kann es seine Aufgabe ausführen.

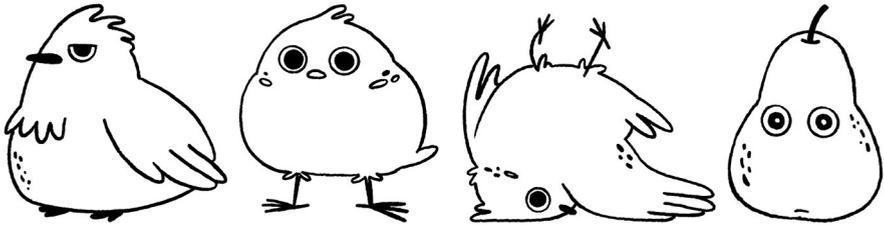


Abbildung 2: Alles Küken?

Da ein Netzwerk nicht von Anfang an weiß, was ein Küken ist, benötigt es eine große Menge an bereits sortierten Bildern, die Küken enthalten. Während der Trainingsphase gewichtet sich das Netzwerk nach den vorgegebenen Bildern. Wenn dem Netzwerk tausend Bilder eines Autos als Küken beigebracht werden, wird es ein Auto als „Küken“ assoziieren. Dieser Lernprozess kann bei Identifizierungen zu automatisierten Vorurteilen führen – wenn das Netzwerk ausschließlich Bilder von Frauen in Kleidern und langen Haaren vorfindet, wird es zunächst nur diese Attribute mit dem Begriff „Frau“ identifizieren. Ein gut trainiertes Netzwerk bedarf eines diversen Trainingssatzes.

Der Output-Layer interpretiert die Werte zu einer Wahrscheinlichkeit, dass das Bild ein Küken beinhaltet. Das Ergebnis wird nun vom Anwender bewertet: Ist es nicht akkurat, werden die Werte erneut gewichtet, bis ein besseres Ergebnis vorliegt. In der Testphase muss das Netzwerk beliebige Bilder als Küken oder kein Küken klassifizieren können.

Für Netzwerke, die Bildinhalte identifizieren sollen, reicht ein betreutes Lernen aus – EntwicklerInnen können die Ergebnisse des Netzwerkes kontrollieren. Wie können nun Bildinhalte *generiert* werden? Wörtliche Beschreibungen besitzen zunächst eine breite Interpretationsfläche, was ihre visuelle Auslegung betrifft. Wenn ich ein Bild von einer Tasse Kaffee sehen will, ist damit nichts über den

⁶ Kompressionsverfahren wie JPEG funktionieren auf ähnliche Weise, indem sie aneinander-grenzende ähnliche Farbtöne erkennen und zu einem Farbton zusammenführen.

Bildstil, die Perspektive oder Art der Tasse gesagt. Allerdings besitze ich gegenüber einem untrainierten Netzwerk eine bildliche Vorstellung, wie das Beschriebene aussehen *könnte*. Nach diesem Prinzip funktioniert das unbetreute Lernen.

Für das unbetreute Lernen werden zwei Netzwerke benötigt, die in einem GAN zusammenarbeiten.⁷ Das Erste Netzwerk spielt die Rolle des Identifizierenden: Es lernt alle Gegenstände und Zusammenhänge, die für das Generieren benötigt werden. Die Trainingsphase eines GANs bedarf bis zu Millionen vorsortierter Bilder. Das Trainieren wird dadurch nicht nur zeitintensiv, sondern verbraucht eine Menge Rechenkapazität, was die Entwicklung eines GANs zu einem kostenintensiven Prozess macht.⁸

Das zweite Netzwerk funktioniert als Gegenspieler des lernenden Netzwerkes. Seine Aufgabe ist es ebenfalls, Bilder zu erstellen, die der Beschreibung entsprechen – nur ohne den Input vorsortierter Bilder. Es startet mit unsortierten Pixeln, die es in eine sinnvolle Ausgabe umwandeln muss. Für den Anfang bekommt das zweite Netzwerk unsortierte Bilder geliefert, die es nachmodelliert.

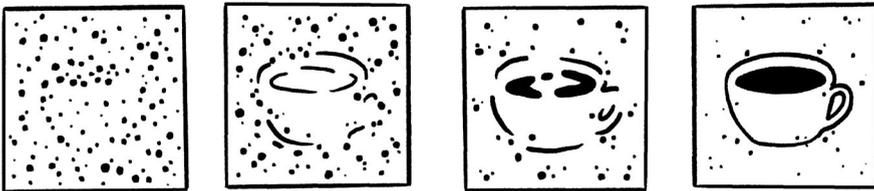


Abbildung 3: Schritte eines GANs während dem Generieren einer Kaffeetasse.

Die Ergebnisse der zwei Netzwerke werden anonym an einen Diskriminator gesendet. Der Diskriminator muss nun beurteilen, welches Bild von dem vortrainierten Netzwerk stammt. Durch das Feedback des Diskriminators nähert sich das zweite Netzwerk den Ergebnissen des ersten an, optimal so weit, dass der Diskriminator nicht mehr zwischen Original und Fälschung unterscheiden kann. Mit dieser Prüfmethode wird die Qualität der generierten Bilder enorm gesteigert. Das unbetreute Lernen kann so ein Netzwerk trainieren, welches eine Vielfalt von

⁷ Vgl. Antonia Creswell et al., „Generative Adversarial Networks: An Overview“, *IEEE signal processing magazine*, 2018.

⁸ Das Trainieren eines Nvidia StyleGANs für eine Bildfläche von 1024x1024 Pixeln benötigt eine Lernzeit von mehr als 41 Tagen, vgl. NVLabs, „StyleGAN – Official TensorFlow Implementation: An Overview“, in [github.com/NVLabs/stylegan] (Zugriff: 28.07.2023).

Bildern kombinieren kann, ohne auf vorgegebene Assoziationen angewiesen zu sein.⁹

Die zunehmende Qualität von GANs führen zu einem Diskurs über den Umgang mit generierter Kunst, Fotos oder Illustrationen. Ende 2022 etablierte sich auf digitalen Plattformen eine „Say no to AI art“ – Bewegung, die die Produktion von generierten Bildern über eine Protestform ablehnt.¹⁰ Neben dem Umgang mit der disruptiven Technologie stellt sich die Frage, wie generierte Bilder in einem visuellen Diskurs funktionieren.

2 Bilder als Diskursmedium

Bilder agieren seit Urzeiten als symbolisches Medium von Kulturen. Ein Abbild steht in der Vermittlung des Gezeigten und des Begriffs. Dabei unterliegt es keiner inhärenten Bedeutung. Eine Bildsemiotik wird durch ständig wandelnde Konventionen, Kontexten und Interpretationspraktiken konstruiert.

Im digitalen Diskurs praktizieren generierte Bilder eine Form des Mimetischen. Mimesis, die „frühe Einheit von [...] physionomischer Darstellung und technognomischer Verfügung“¹¹, versteht das Erkennen von Sachverhalten als Dialektik des geistig Erfahrbaren und des materiellen Auslösers. Die Intention des Anwendenden kann durch GANs zu einer visuellen Erfahrung generiert werden. Das Kriterium der Repräsentation wird nur nachgeahmt. Generierte Bilder stehen nicht in der vollen Autorschaft des Anwendenden und sind auch kein Abbild der Realität. Allerdings führen sie in ihrer Symbolik eine bewusste Konfrontation sowohl gegen die Autorschaft als auch gegen den Anspruch auf einen Wahrheitsgehalt.

Ein Beispiel dafür ist das generierte Bild von Papst Franziskus in einer modischen Daunenjacke, das Anfang 2023 einen Punkt in der Debatte markierte.¹² Der Anwender Pablo Xavier, ein Bauarbeiter aus Chicago, veröffentlichte das Bild zunächst als Meme im Internet. Die Wirkung des Komischen führte mit seiner herausragenden Qualität jedoch zu Desinformationen über den Papst, Kritik an der

⁹ Vgl. Scott Reed et al., „Generative Adversarial Text to Image Synthesis“, *PLMR*, Bd. 48, 2016, 1060–1069.

¹⁰ Vgl. Oliver Bunte, „Artstation: Künstler protestieren gegen KI-Kunstwerke und KI-Trainings“, *heise*, 2022, in [heise.de/news/Artstation-Kuenstler-protestieren-gegen-KI-Kunstwerke-7397005.html] (Zugriff: 28.07.2023).

¹¹ Jürgen Habermas, „Anthropologie“, in *Das Fischer Lexikon Philosophie*, hg. von Alwin Diemer/Ivo Frenzel, Frankfurt a. M. 1958, 18–35, 28.

¹² Vgl. Christian Schicha, „KI-generierte Deep Fakes. Wie gehen wir mit einem Papst im Daunenmantel um?“, *deutschlandfunk*, 2023, in [https://t1p.de/e0jh4] (Zugriff: 28.07.2023).

katholischen Kirche und vermehrter Unsicherheit über die Bebilderung von Berichten.



Abbildung 4: Ceci n'est pas un pape.

Der Fall Xaviers veranschaulicht, wie ein generiertes Bild zwar kein direktes Abbild darstellt, es aber imitiert und als solches eingesetzt werden kann. Im Gegensatz zur Fälschung repräsentiert das generierte Bild keine Referenz. Es wird zum eigenständigen Zeichen, dessen Bedeutung und Interpretation nicht eindeutig sind.

Der Kontrast zwischen dem generierten Bild als nicht-referentielles Abbild und der Verwendung von GANs zur Stilreproduktion verdeutlicht die zunehmende Komplexität in der Beziehung zwischen visueller Repräsentation und künstlerischer Intention. Während das generierte Bild in seiner Semantik variabel ist, wird seine Stilistik durch GANs wie Midjourney in einem strukturierten, reproduzierbaren Rahmen dargestellt. Mit trainierten GANs können dabei nur Stile reproduziert werden, die in großen Datensätzen prominent sind. Stile, von denen es wenige Exemplare gibt, sind auch schwieriger zu trainieren. Große GANs sind also keine Universalmaschinen, die alles reproduzieren können. Einerseits ermöglicht die Technologie neue Formen der künstlerischen Äußerung durch das Prompten von vorher unzugänglichen Stilen, andererseits entstehen durch die Automatisierung des Stils neue Begrenzungen für die bildliche Darstellung. Was bedeutet das für die Illustration?

2.1 Illustration im Paradigmenwechsel

Der Begriff der Illustration steht in einem Spannungsfeld zwischen den Disziplinen der Kunst und des Designs. Illustration besitzt das Ziel einer Inhaltsvermittlung und

agiert zugleich im Stil einer Autorschaft. Was wird dargestellt – und wie? Durch ihre visuelle Kommunikation entsteht eine Wissenstransformation. Der zu vermittelnde Inhalt wird mithilfe einer semiotischen Logik zugänglicher. Gedankengänge können durch eine Ausführung illustriert oder Illustration als die Bebilderung eines Textes verstanden werden. Grundlegend beschreibt Illustration die Darstellung einer intentionalen Idee, traditionell mit zeichnerischen Mitteln.

In einer zunehmend bebilderten Welt befindet sich die Illustration in einer ubiquitären Position. Otto Neurath beschrieb Anfang des 20. Jahrhunderts die gesellschaftliche Entwicklung, von Bildern zu lernen und durch sie zu kommunizieren, als „großen Teil ihres Wissens und ihrer allgemeinen Bildung durch bildhafte Eindrücke“.¹³ Die Transformation zu einem „visuellen Zeitalter“¹⁴ bedeutet für die Illustration eine erweiterte Anwendungsfläche sowie eine Verantwortung gegenüber ihrer Semiotik.

Wo eine Illustration eingesetzt wird, ist von ihrem Anwendungsfeld abhängig. Kein Beispiel lässt sich auf eine feststehende Anwendung reduzieren. Allerdings existiert für IllustratorInnen vorher ein Konsens, wie etwas dargestellt wird.

Die formelle Illustration versucht, das Abzubildende zu erklären. Für eine neutrale Darstellung orientiert sie sich an Gestaltungsregeln, die zu einer größtmöglichen Lesbarkeit beitragen. Baupläne, die biologische Illustration einer Katze oder das Schriftsystem „Isotype“ vermitteln einen Lerngehalt, der durch eine möglichst lesbare Stilistik zugänglich wird.

Werbende Illustrationen passen ihre Bildsprache an eine Zielgruppe an. Sie versuchen, ihr Publikum von einer dargestellten Idee zu überzeugen. Der Autor passt sich an das Idealbild des Auftraggebers an, während die Stilistik des Autors zur dessen visuelle Identität übernimmt. Gleichzeitig wirkt die Illustration als Ästhetisierung des Dargestellten, die eine repräsentative Werbefigur bildet. Der US-amerikanische Illustrator Gil Elvgren verwendete für Coca-Cola Szenarien eines Pin-Up-Girls. Es entsteht eine Symbolik, die das Sehen der Dame mit dem Produkt assoziiert. Die werbende Illustration funktioniert so als ästhetische Manipulation.

Subjektive Illustrationen kommunizieren aus der Perspektive der eigenen Autorschaft. Sie vermittelt einen phänomenologischen Bildgehalt: Wie nehme ich etwas wahr? Die Illustration eines Teddybärs in Pastellfarben und rundem Strich wirkt anders als ein kubistischer Stil in Grautönen. Für welche Stilistik sich entschieden wird, liegt an den ästhetischen Vorhaben des Autors, die sowohl individuell als auch kulturell bedingt sind.

¹³ Otto Neurath, „Statistische Hieroglyphen, in *Diagrammatik-Reader: Grundlegende Texte aus Theorie und Geschichte*, hg. von Birgit Schneider/Christoph Ernst/Jan Wöpping, Berlin/Boston 2016 [1927], 213–218, 213.

¹⁴ Vgl. Gerhard Paul, *Das visuelle Zeitalter. Punkt und Pixel*, Göttingen 2016, 712.

Die untenstehende Illustration ist die klassische Bebilderung einer Idee. Sie hilft, das Beschriebene besser zu verstehen. Illustratoren als „Maler des modernen Lebens“, wie sie Charles Baudelaire beschrieb, äußern über den Stil ihre Wahrnehmung, während sie gleichzeitig über das Abgebildete informieren.¹⁵

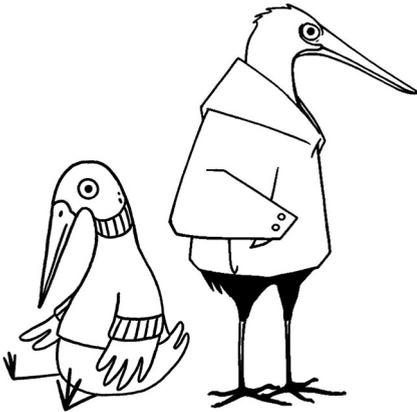


Abbildung 5: Die Comicfigur „Härbärt“^(a), links im Original, rechts eine generierte Illustration im verwandten Stil, erzeugt mit Midjourneys Niji 5.0.^(b)Vgl. Alisa Geiß, „HÄRBÄRT“, Webcomic, 2022, in [alisageiss.com/haerbaert-comic] (Zugriff: 28. 07. 2023).

Das zentrale Paradigma der Illustration ist das Skizzieren. Im Gegensatz zum Prompting erfolgt das Skizzieren einer Idee nicht über eine detaillierte Beschreibung, sondern über das Zusammenspiel der bildlichen Vorstellung und der Beurteilung ihres Abbilds, der Skizze. Skizzieren ist ein explorativer Prozess, in dem der Freiraum, mit Form und Darstellung zu experimentieren, einen Lerngehalt über die eigene Wahrnehmung beinhaltet.

Prompting unterscheidet sich fundamental von diesem Ansatz. Über Prompts, detaillierte Beschreibungen des gewünschten Bildes, können die trainierten Stile eines GANs dem beschriebenen Bild zugeordnet werden. Während das Skizzieren einen unfertigen Prozess darstellt, kombiniert das Prompting die Vorstellung mit einem ausgearbeiteten Ergebnis. Hierbei nimmt die Ebene des Textes eine ähnliche Rolle des Skizzierens ein.

¹⁵ Vgl. Charles Baudelaire, zitiert nach Andreas Rauth, „Unverwechselbar. Über Stil und Authentizität in der Illustration“, in *superILLU. Zu einer Theorie der Illustration*, hg. von Juliane Wenzl et al., 2022, 197–205, 197.

2.2 Das Verhältnis von Text und Bild zu Prompts

Beim Verfassen von Prompts ist eine präzise Beschreibung entscheidend. Was nicht beschrieben wird, kann auch nicht generiert werden – der Prompt einer „Ente“ spezifiziert nicht, ob damit die Tierart oder das Automodell gemeint ist. Ein bildliches Denken hilft auch beim Prompten. Die spezifischen Bedingungen der jeweiligen Bildelemente müssen erkannt und beschrieben werden. Die Wortwahl des Prompts beeinflusst ebenfalls das Ergebnis. Ein Prompt, der auf eine konzeptionelle Darstellung abzielt, benötigt eine andere Formulierung als ein Prompt, der ein Bild exakt reproduzieren will.



Abbildung 6: Prompt: „a chunky round white dove wearing a tiny small black hat, carrying a black briefcase, not colored, just outlines“, erzeugt mit Midjourneys Niji 5.0.

Das Erstellen von Prompts über einen Textgenerator ermöglicht eine automatisierte Herstellung von Prompts. Dies bewirkt einen verbesserten Output innerhalb GANs, da das Anpassen eines Prompts ein zeitintensiver Prozess sein kann und nach einem Trial-and-Error-Prinzip funktioniert. Durch eine Automatisierung kann die Reichweite an möglichen Formulierungen erhöht werden, womit die Grenzen eines GANs ausgetestet werden können.

Ist das „Prompten“ also als Fähigkeit wahrzunehmen? Mit einem Rückblick auf die Erfindung der Fotografie wurde diese von Baudelaire stark kritisiert. Fotografierende seien „gescheiterte Maler mit zu wenig Talent“¹⁶. Baudelaire argumentierte, dass Fotografie lediglich eine Entwertung der menschlichen Vorstellungskraft darstelle. Seine überholte Position erinnert an das Argument, generative KI sei für das kreative Schaffen kontraproduktiv.

¹⁶ Vgl. Charles Baudelaire, zitiert nach James Rubin, *Impressionism and the Modern Landscape*, Berkeley 2008, 43–45.

Im Kontrast zur Fotografie besitzen GANs kein klares Verhältnis zwischen Werk und Schöpfenden. Ein Gedankenexperiment:¹⁷ Eine Kunstsammlerin gibt einem menschlichen Künstler, den sie nicht kennt, einen Zettel mit ihrer Idee für ein Kunstwerk. Der Künstler setzt diese Idee um und übergibt ihr das fertige Werk. Die Kunstsammlerin ist trotz ihrer originellen Idee für das Werk nicht die Schöpferin, sondern der Künstler. Bei KI verhält es sich zunächst ähnlich: Auch wenn jemand eine Idee als Prompt formuliert und KI ein Ergebnis produziert, hat der Anwendende nicht das Werk erschaffen. Da KI jedoch nicht als Person gilt, liegt die Erschaffung des Werks weder auf der Seite des Anwendenden noch auf der Seite der Algorithmen. Das Prinzip der Autorschaft scheint bei generativen Prozessen nicht bestimmbar zu sein.¹⁸ Damit wird ein Problem sichtbar, das Roland Barthes' literarischem „Tod des Autors“ nahekommt: Führt die Abwesenheit des Autors zu einer obsoleten „Entzifferung“ des Arbeitsprozesses?¹⁹

2.3 Autorschaft und (Bild-)Generatoren

Der Diskurs über den Stellenwert von GANs ist jung und durch ihre wachsende Verfügbarkeit angetrieben. Positionen reichen von der Ablehnung von GANs und Klagen zum (vermeintlichen) Missbrauch des Urheberrechts²⁰ bis zum KI-Futurismus um Raymond Kurzweil und der Marginalisierung von Kritik.²¹

Die Initiative von Illustratoren, die Existenz von GANs zu ignorieren, ist eine Protestreaktion. Sie bietet keine nachhaltige Verteidigung des gestalterischen Handwerks gegenüber KI. Die eigentliche Botschaft ist die Erkenntnis, dass die eigene Arbeit entfremdet und automatisiert werden kann. Dabei ersetzt die Präsenz von GANs nicht die Rolle der kreativ Tätigen. Was stattfindet ist eine Neuorientie-

¹⁷ Vgl. Chris Shehan, 31.08.2022, in [twitter.com/ChrisShehanArt/status/1564821938400264192] (Zugriff: 27.08.2023).

¹⁸ Für weiterführende Betrachtungen zur Autorschaft und Urheberrecht verweise ich auf den Tagungsbeitrag von Amrei Bahr.

¹⁹ Vgl. Roland Barthes, „Der Tod des Autors“, in *Das Rauschen der Sprache*, Frankfurt a. M. 2006 [1969], 57–63, 62.

²⁰ Vgl. Initiative Urheberrecht, „Urheber:innen und Künstler:innen fordern Maßnahmen zum Schutz vor generativer KI in der Europäischen KI-Verordnung“, Stellungnahme 14.09.2023, in [https://t1p.de/m6nab] (Zugriff 04.10.2023)

²¹ Vgl. Raymond Kurzweil, *The Singularity Is Near: Why Humans Transcend Biology*, New York 2005 oder Anudeex Shetty/Nivesh Raj, „A Study on Recent Advances in Artificial Intelligence and Future Prospects of Attaining Superintelligence“, in *Proceedings of Third International Conference on Communication, Computing and Electronics Systems*, Bd. 844, hg. von V. Bindhu/João Manuel R. S. Tavares/Ke-Lin Du, Singapur 2022, 879–892.

nung ihrer Arbeitsweise. Die Kamera hat den Maler nicht ersetzt, sondern ihn von seinen Portraitaufträgen befreit. Die Kontroverse liegt in den sozialwirtschaftlichen Faktoren, die die Arbeit von Kreativen erst ermöglicht und nun durch eine Disruption zu verringern drohen.

Der im Jahr 2023 begonnene und erst im Herbst beendete *Writers Guild*-Streik in Hollywood zeigt die Bereitschaft von Unternehmen, generierte Inhalte in die Kulturwirtschaft zu integrieren. Das dreijährige Abkommen zwischen Autoren und diversen Netzwerken verhandelt den Umgang mit KI zugunsten der Autoren in der Schreibbranche. Zunächst besteht bei der Arbeit mit KI eine Kennzeichnungspflicht.²² Dies soll vor einem Vergleich zwischen menschlichen und generierten Schriften schützen und die Integrität des Autors bewahren. Das Schreiben mit KI darf zusätzlich nicht als Arbeitsmethode vorausgesetzt werden.

Als Kernelement ist in der Vereinbarung unter Artikel 72 B deklariert, dass KI keine Person darstellt und damit keinen Anspruch auf Urheberrechte besitzt oder in einer Form entlohnt werden muss.²³ Amerikanische Autoren behalten ihre Urheberstellung, wenn sie mit generierten Inhalten arbeiten. Sollte das Unternehmen einem Autor generiertes Material zur Verfügung stellen, wird dieses als literarisches Material betrachtet und interferiert nicht mit der Autorschaft.²⁴ Das Unternehmen darf das erstellte Material zudem nicht veröffentlichen oder nutzen, um diese Vereinbarung zu umgehen.

Ein Problemfeld für Autoren bleibt das unmögliche Überprüfen der Datensätze von Text- und Bildgeneratoren auf potenzielle Urheberrechtsverstöße. Im deutschen Raum ist die Inhaltserstellung über Data Mining seit dem 2021 erlassenen § 44b Urheberrechtsgesetz (UrhG) definiert. Jedoch ist im Gegensatz zum existierenden § 60d UrhG, das sich auf Text- und Data Mining in der Wissenschaft bezieht, mit § 44b die Verarbeitung von Daten auch im privaten Raum gestattet.²⁵ Die Erlassungen werden mit einer Anregung von Innovation in der Privatwirtschaft begründet.

Obwohl die Reproduktion von urheberrechtlich geschützten Inhalten wie Comicfiguren ohne die Einwilligung des Urhebers dank §§ 19a und 44 UrhG weiterhin geschützt ist, bleibt die Veröffentlichung des Werks die Voraussetzung für die Integration in Generatoren. Eine Grauzone bildet das *Web Scraping*, dem Sammeln von repräsentativen Daten über Suchmaschinen oder ähnliche Websites. Exem-

22 Vgl. „Memorandum of Agreement for the 2023 WGA Theatrical and Television Basic Agreement“, in [wga.org/uploadedfiles/contracts/2023_mba_moa.pdf] (Zugriff: 26.09.2023), 69.

23 Ebd., 68.

24 Ebd., 69.

25 Vgl. Benjamin Raue, „Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG)“, *Zeitschrift für Urheber- und Medienrecht*, Baden-Baden 2021, 793–802.

plarisch genügt das Abbild eines Gemäldes in einem Online-Shop, welches über eine Bildersuche gefunden werden kann. Darunter leiden besonders Künstler und Autoren, die eine weite digitale Präsenz besitzen. Bildgeneratoren wie Midjourney oder StableDiffusion reproduzieren unter der Nennung bestimmter Autoren gerne mal ein Wasserzeichen. Ob KI-Unternehmen bei Verstößen von generiertem Output auch haften müssen, bleibt ungeklärt und von den Nutzungsbedingungen des Anbieters abhängig. So heißt es zum Beispiel bei Midjourney:

We provide the service as is, and we make no promises or guarantees about it. [...] If You [sic] knowingly infringe someone else's intellectual property, and that costs us money, we're going to come find You and collect that money from You. We might also do other stuff, like try to get a court to make You pay our attorney's fees. Don't do it.²⁶

Midjourney-CEO David Holz beschreibt den Bestand seines Datensatzes als „große Repräsentation des Internets“.²⁷ Das Generieren eines Prompts basiert auf dem Erfahrungsgehalt von Millionen abrufbaren Bildern. Generative Algorithmen stützen sich auf die „vergegenständlichte Arbeit“²⁸ eines Menschen, dessen Werke im Internet veröffentlicht wurden. Der Schematismus der Bilder wird über die Maschine in eine Dienstleistung integriert; ein Prozess, den Marx als „general intellect“ beschrieb:

Die Entwicklung des *capital fixe* zeigt an, bis zu welchem Grade das allgemeine gesellschaftliche Wissen, knowledge, zur *unmittelbaren Produktivkraft* geworden ist und daher die Bedingungen des gesellschaftlichen Lebensprozesses selbst unter die Kontrolle des *general intellect* gekommen und ihm gemäß umgeschaffen sind.²⁹ Bei Marx wird unter dem allgemein zugänglichen Wissen das verstanden, was eine vorangegangene Generation der Jetzigen zur Verfügung stellt.³⁰ Niemand muss den Computer neu erfinden, um auf sein Konzept zugreifen zu können. Genauso sind Handwerke, physikalische Gesetze oder Weltkarten Produkte von vergesellschaftetem Wissen, welches“ nicht neu erarbeitet werden muss.

²⁶ Midjourney, „Terms of Service“, Abs. 10 „Limitation of Liability and Indemnity“, in [docs.midjourney.com/docs/terms-of-service] (Zugriff: 25.09.2023).

²⁷ Vgl. David Holz, „Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy“, *Forbes*, in [forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy] (Zugriff: 28.07.2023).

²⁸ Karl Marx, „Der relative Mehrwert“, in *MEW*, Bd. 43, Berlin 1983 [1858], 321.

²⁹ Karl Marx, „Grundrisse der Kritik der Politischen Ökonomie“, in *MEW*, Bd. 42, Berlin 1983 [1858], 602, Hervorh. im Original.

³⁰ Vgl. Wolfgang Fritz Haug, „General intellect“, in *Historisch-kritisches Wörterbuch des Marxismus (HKWM)*, Bd. 5, hg. von Institut für Kritische Theorie (INKRIT), Hamburg 2001, 230–242.

Mit der Automatisierung des Arbeitsprozesses wird die Expertise akkumuliert. Die Maschine löst als effizientere Produktivkraft den Arbeiter ab, dieser tritt „neben den Produktionsprozess, statt sein Hauptagent zu sein.“³¹ Die Verbreitung und Nutzung des öffentlichen Wissens stehen dabei in Konkurrenz mit der Arbeitszeit. Ein Grund für Unternehmen, ihre auf Open-Source gebaute Software zu privatisieren.

Das Silicon Valley, darunter OpenAI als aktueller Hauptakteur der KI-Entwicklung, schildert die Akkumulation von Expertise mit einem dualistischen Narrativ.³² Einerseits schaffe der technologische Fortschritt Innovationen, die in allen Lebensbereichen Disposition finden und die alltägliche Arbeit erleichtern. Andererseits ist den Entwicklern bewusst, dass der intentionale Missbrauch der Technik nur schwer beherrschbar ist. Die Ambivalenz des Fortschritts- und Destruktionsgedankens vollzieht sich innerhalb eines paradoxen Systems: Die Korrelation zwischen Technik als Hilfsmittel des gesellschaftlichen Fortschritts und postfordistischer Wirtschaft, die die Digitalisierung als zwingende Voraussetzung ihres Wachstums benötigt, zeichnet eine geiselhafte Existenz der KI.

Im gesellschaftlichen Diskurs manifestieren sich zwei Aspekte, in denen die ambivalente Darstellung von KI verschiedene Mythen befördert. Im Folgenden werden sie erläutert.

3 Der Mythos als Immanenz des Generativen

Erstens befördert die mangelnde Nachvollziehbarkeit der Wahrscheinlichkeit sowie die fehlende Auskunft über den Inhalt der Datensätze die Wahrnehmung von KI-Generatoren als eine eigentumslose Quelle. Das Resultat simuliert ein „Vorfinden“ des Eingebenen. Dabei beeinflusst die hohe Qualität der Ergebnisse die Auffassung des jeweiligen Mediums, wodurch das Unterscheiden von realen und generierten Inhalten an seine Grenzen gelangt. Die Integration ähnelt einem „Kunstmythos“, der sich den grundlegenden Elementen des Mythos bedient.³³

Hans Blumenberg definiert Mythen als „Geschichten von hochgradiger Beständigkeit ihres narrativen Kerns und ebenso ausgeprägter marginaler Variationsfähigkeit“.³⁴ Sie schützen das menschliche Bewusstsein vor seiner Auslieferung

31 Vgl. Marx 1983, 601.

32 Vgl. Jakob von Lindern/Jochen Wegner, „Sam Altman: ‚Irgendwann werden wir ein Modell erschaffen, das gefährlich ist‘“, *Zeit Online*, 26.05.2023, in [zeit.de/digital/2023-05/sam-altmann-openai-ceo-chat-gpt-ki] (Zugriff: 04.10.2023).

33 Vgl. Hans Blumenberg, *Arbeit am Mythos*, Frankfurt a. M. 1961, 194.

34 Ebd., 40.

in eine unvorhersehbare Welt – allerdings passiert dies nicht über ein makellostes Bildnis. Was einen Mythos glaubwürdig macht, ist die Übertragbarkeit seines dialektischen Erfahrungsgehalts.³⁵ Die Grundprinzipien von Mythen sind so elementar, „daß sie immer wieder überzeugen, sich immer noch als brauchbarster Stoff für jede Suche nach elementaren Sachverhalten des menschlichen Daseins anbieten.“³⁶

Eine Blumbersgsche Perspektive auf die Relation zwischen Mythos und Logos zeigt, dass der Mythos nicht als Antonym zum Logischen verstehen ist. Der vermeintliche Übergang des Mythischen zu einer rationalen Epistemologie übersieht den Aufbau und Grund des Mythos. Dessen logischer Kern besteht in einem Schutzmanöver vor der chaotischen Wirklichkeit, die keine Struktur verfolgt.³⁷

Die Remythisierung, das Wiederaufleben eines Grundmythos unter zeitgenössischen Phänomenen, ist ein Produkt abwesender Sinnhaftigkeit. Sie trägt den Nutzen einer Gegenwartsbewältigung: In der Existenz einer nachmythischen Epoche sucht sie nach neuen Machtverhältnissen, da die alten Mythen nicht auf die Gegenwart übertragbar sind.³⁸ Dieser Anspruch auf Beständigkeit ist besonders für ungewisse Zeiten attraktiv:

Das macht Zeiten mit hohen Veränderungsgeschwindigkeiten ihrer Systemzustände begierig auf neue Mythen, auf Remythisierungen, aber auch ungeeignet, ihnen zu geben, was sie begehren. Denn nichts gestattet ihnen zu glauben, was sie gern glauben möchten, die Welt sei schon immer so oder schon einmal so gewesen, wie sie jetzt zu werden verspricht oder droht.³⁹

Eine Remythisierung kann in ihrem Wahrheitsgehalt nicht affirmiert werden, obwohl sie Anspruch auf ihre Gültigkeit erhebt. Sie projiziert eine Sinnfindung, die sich in ihrer Verantwortung gegenüber der Realität als zeitlos inszeniert und somit entzieht. Als Schutzsphäre nimmt sie die Angst vor der eigenen Entscheidungsfähigkeit.⁴⁰ „Die Unfähigkeit, Substitutionen vorzunehmen oder gelten zu lassen, ist nahezu identisch mit der anderen, Delegation von Zuständigkeiten an andere vorzunehmen und Repräsentationen für Entscheidungen der Vielen durch Wenige gelten zu lassen.“⁴¹

Im Mythos der KI finden sich die abwesende Sinnhaftigkeit sowie eine Schwierigkeit in der Bestimmung ihres Wahrheitsgehalts im Potential wieder,

35 Vgl. ebd., 33–34.

36 Ebd., 166.

37 Ebd., 18–20.

38 Vgl. ebd., 178.

39 Ebd., 41.

40 Vgl. ebd., 178.

41 Ebd., 13.

hochwertige Desinformation zu verbreiten. Obwohl einzelne Inhalte auf ihren Wahrheitsgehalt überprüft werden können, verändert die Integration von generierten Inhalten auch die Kommunikation innerhalb der digitalen Sphäre.

Der unreflektierte Glaube an eine technische Beschaffenheit ohne die Berücksichtigung ihrer Anwendenden, der Missbrauch der instrumentellen Vernunft, wie Horkheimer sie beschrieb, nutzt Technik als Begehrensubjekt der Pragmatik.⁴² Der „Triumph des Mittels über den Zweck“⁴³ führt zur Konsequenz einer absoluten Gewichtung des Mittels, welche in ihrer Funktionsweise weder zweck- noch werterational ist und eine eigene, glaubwürdige Instanz formt. Diese kann „zu einer Grundlage [...] werden, welche Kultur ausdrückt und reguliert“.⁴⁴

Bedeutet das also, dass Technik als dogmatischer Pragmatismus schon immer einem Selbstzweck verfällt? Freilich nicht. Ihr Sinn ist es nicht, destruktiv auf die Natur und den Menschen zu wirken, um in einer informationsästhetischen Barbarei zu enden. Technologie ist ein Handwerk, das gebraucht und missbraucht werden kann. Technischer „Fortschritt“ bedeutet eine verbesserte Funktion bereits eingesetzter Technik zu dem Punkt, wo der Austausch von Techniken eine Revolution des ökonomischen Systems bewirkt und damit auch verbesserte Lebensbedingungen erzielt.

Zweitens erfordern Prompts wörtliche Beschreibungen, die für die reine visuelle Erkenntnis nicht erforderlich waren. Dies besitzt Auswirkungen auf die Assoziationen von Text und Bild. Wenn generierte Inhalte Formulierungen zugeordnet werden, die keinen logischen Anspruch erheben, besitzt die KI weiterhin die Assoziation mit jener Formulierung. Es steht zur Frage, inwieweit Prompts die Konstruktion von Bedeutungen beeinflussen werden.

Prompts bilden die Schnittstelle zwischen KI und Mensch. Der Vergleich menschlicher Fähigkeiten mit der einer KI ruft eine bestimmte Vorstellung im Geiste des Empfängers hervor. Sowohl komplexe als auch einfache Sachverhalte können durch Metaphern verständlicher dargestellt werden. Ob diese Tendenz unausweichlich ist, bleibt offen; feststeht, dass der Gebrauch davon zu differenzierten Vorstellungen führen kann. Die Aussage: „Eine KI denkt wie ein Mensch“, löst unterschiedliche Vorstellungen aus, was das zu bedeuten hat. Beherrscht sie Kreativität? Logik? Macht sie Fehler, hat sie Vorurteile, denkt sie überhaupt? Der Vergleich mit den menschlichen Fähigkeiten bewirkt eine Anthropomorphisierung

⁴² Vgl. Max Horkheimer, „Zur Kritik der instrumentellen Vernunft“, in ders., *Gesammelte Schriften*, Bd. 6, hg. von Alfred Schmidt, Frankfurt a. M. 1991 [1947/1967], 19–186, 113.

⁴³ Ebd., 60–61.

⁴⁴ Gilbert Simondon, zitiert nach Jean-Hugues Barthélémy, „Simondon – Ein Denken der Technik im Dialog mit der Kybernetik“, in *Die technologische Bedingung. Beiträge zur Beschreibung der technischen Welt*, hg. von Erich Hörl, Berlin 2011, 93–109, 95.

der KI. Sie löst Vertrautheit über ihre Funktionsweisen aus, welches zu einer Mythisierung des Verhältnisses von Mensch und Maschine beiträgt.

Wie schwierig es ist, einen Mythos zu beenden, zeigt sich in seinem metaphorischen Zirkel, der nach dem Ableben eines Mythos brauchbar wird. Der Mythos verkörpert ein gesellschaftliches Bewusstsein, das von den Erfahrungen seiner Umwelt abstrahiert. Sein Lehrgehalt besteht in einer reflektierten Rezeption seiner Entstehung. So unmöglich das Verschwinden von mythischen Erzählungen aus der Welt ist, so lehrreich ist der Mythos in einer Deutung, die zur Progression der Gegenwart beiträgt.⁴⁵

Schließlich besitzen Generatoren im Kontext ihrer Unbestimmtheit das Potential zur Dekonstruktion traditioneller Entwurfspraktiken. Während bürokratisierte Prozesse oft von normativen Strukturen geprägt sind, bieten Textgeneratoren die Möglichkeit, jenseits dieser Konventionen zu agieren. Dies impliziert eine Neubewertung des Entwerfens als solches und kann langfristig zu freieren Text- und Bildpraktiken führen. Wenn Bildgeneratoren vermehrt für Stockfotos eingesetzt werden, wenn während dem Schreiben nicht behauptet werden kann, dass für sich selbst geschrieben wurde, so ist das vielleicht der Beginn eines Verständnisses, dass Generatoren standardisierte Produktionen entlasten können. Eine Chance von KI ist, den Mediengebrauch aus seiner Bürokratisierung zu befreien.

45 Blumenberg 1961, 681.

