KUNST
CHRONIK

Sites of Action and Reflection

Derivate.
Zum Reproduktionsbegriff synthetisch generierter Bilder

Dr. Francis Hunger
Akademie der Bildenden Künste München
francis.hunger@adbk.mhn.de

https://doi.org/10.11588/kc.2025.7.111292

78. Jahrgang | Ausgabe 7 | Juli 2025

Derivate. Zum Reproduktionsbegriff synthetisch generierter Bilder

Francis Hunger

Die aus den mimetischen Diensten entlassenen generierten Bilder verkommen, so der Kunsthistoriker Wolfgang Ullrich, zu „Bilderbullshit" (Ullrich 2025), denn der „Wahrheitsgehalt von Bildern interessiert nicht mehr". Dafür wird zunehmend der Begriff „AI Slop", deutsch in etwa: „KI-Mansche", verwendet (Koebler 2025). Wenn, wie von Ullrich angedeutet, sich die Referenz von Originalgegenstand und Abbild auflöst und wenn Ästhetik durch Mansche ersetzt wird, stellt sich die Frage nach der Reproduktion, einem wesentlichen Begriff der Kunsttheorie, neu.

Der Text fokussiert auf zwei Argumente: die derivative Ableitung synthetischer Bildsujets aus den zugrundeliegenden Basiswerten einerseits, und den Verlust des indexikalischen Bildes zugunsten eines korrelativen Bildes andererseits. Die Software-Verfahren zur Bildgenerierung sind nicht mit Vergleichen zu künstlerischen Werkzeugen zu fassen. *Stable Diffusion* oder *MidJourney* sind kein Zirkel, Meißel, Pinsel, keine Linse. Sie werden von keiner Hand geführt, benötigen kein Körpergedächtnis und sind kein Werkzeug im klassischen Sinne. Synthetische KI-Bildgeneratoren sind am ehesten als parametrische Maschinen (Pasquinelli 2023, 49) zu fassen, beziehungsweise im Sinne der Infrastructure Studies (Bowker u. a. 2010) als komplexe Infrastrukturen, wie sie beispielsweise Kate Crawford und Vladan Joler in der Kartierung *Anatomy of an AI System* (2018) diskutiert haben. Häufig ist zu beobachten, dass der Begriff „Künstliche Intelligenz" und das Verwenden anderer anthropomorphisierender Termini wie „sehen" und „lernen" anzeigen, dass durch die Autor*innen vor einer phantasmatischen Folie argumentiert wird, die den statistischen Verfahren menschliche Agency zusprechen will (Hunger 2023). Dem ist eine vertiefte konzeptuell-mediale Auseinandersetzung entgegenzusetzen, inklusive der Fähigkeit der Autor*innen, das Feld generativer und synthetischer Medien mittels Programmierkenntnissen zu durchdringen.

Was stellen diese parametrischen Maschinen, die salopp „Künstliche Intelligenz" genannt werden, her? Die synthetischen Pixelassemblagen, die mittels generativer KI erzeugt werden, sind flach, sie enthalten keine Perspektive, keinen Fluchtpunkt und keinen Sehpunkt, sie verlachen die *Constructio*. Damit stellt sich die Frage nach dem Standpunkt, dem Ort der Betrachtung durch menschliche Betrachter*innen und nach menschlicher Subjektivität neu. Zwar bildete das klassische Gemälde oder die Fotografie „Realität" nicht eins zu eins ab, sondern war eine zweidimensionale Übertragung von künstlerischen Ideen, Interpretationen, Übersetzungen von Wirklichkeit. Jedoch folgten diese einer Mimesis-Ästhetik, welche bei synthetischen Bildgenerierungsverfahren aufgekündigt wird.

Was ist Reproduktion?

Um diesen kategorialen Neujustierungen nachzugehen, umreiße ich im Folgenden den Reproduktionsbegriff, diskutiere die Datengrundlage synthetischer Bilder und beschreibe die dazu nötigen, als ‚Künstliche Intelligenz' bezeichneten Softwareverfahren der Transformer und Diffusionsnetzwerke. Nach dieser operationalen Betrachtung diskutiere ich, wie synthetischer Bildoutput ästhetisch argumentiert. Über das Argument der Verflächigung und Optizität verdeutlicht der Text, wie sich diese Verfahren von bisherigen Bildreproduktionsverfahren unterscheiden, um schließlich mit der Metapher der (Finanz-)Derivate den Reproduktionsbegriff neu aufzustellen.

Reproduktion wird typischerweise als die möglichst genaue, skalierte Wiedergabe eines Objektes mittels

eines anderen, meist transportablen Mediums verstanden. Über Jahrhunderte haben Menschen kulturtechnisch erlernt, diese Reproduktionen zu lesen und, in ästhetisierte Gebrauchsgegenstände verwandelt, zu genießen und zu nutzen. In der technischen Reproduktion konzentriert sich der Diskurs auf die möglichst detailgetreue Darstellung, welche in Abwesenheit des Originalgegenstandes idealerweise seine Rekonstruktion ermöglichen würde (vgl. die Beiträge von Siegel 2020 und Altinoba 2020). In der künstlerischen ‚Reproduktion' hingegen ist eine gewisse Differenz zwischen Ursprungsobjekt und Darstellungsweise sogar gewünscht, denn sie eröffnet den symbolischen Raum, der dem bürgerlichen Subjekt die schambefreite Erfahrung des *plaisir* (Barthes 1973, 22–24) beim Betrachten von Kunst ermöglicht. Gegenüber dem Original wird die Reproduktion als defizitär positioniert, da sie eben nicht das Original sei.

Die so zirkulierenden Kunstwerke wiederum behaupten Originalität, gekoppelt an die Subjektivität der in der Moderne als autonom konzipierten Kunst. Deren massenhafte Reproduktion wurde maßgeblich durch Walter Benjamin als Angriff auf die Aura des Originals beschrieben. Benjamin löste für die industrielle Gesellschaft das Auratische kritisch in Richtung von Massenmedien und Berühmtheit von Stars auf, welche heute in ihrer destillierten Form als Influencer und Realityshow-Star allgegenwärtig ist (Benjamin [1935] 1991, 452). Gleichzeitig warnte er vor den faschistoiden Potentialen massenmedialer Bildproduktion – in seiner Zeit vergeblich. In heutigen generativbildgebenden Medien begegnet uns dieses Gespenst im Dienst rassistischer, sexistischer und regressiver Bildsprachen erneut (Meyer 2025). Appropriation Art von Künstler*innen wie Marcel Duchamp, Richard Prince und Elaine Sturtevant deklinierten Original, Kopie und Reproduktion für die Postmoderne.

Mit den generativen KI-Verfahren sind die Betrachter*innen nicht mehr mit Reproduktionen konfrontiert, die sich auf einzelne und als solche identifizierbare Originale beziehen, die sorgfältige Aufmerksamkeit erfahren, sondern mit der massenhaften Generierung von Bilderzeugnissen, extrahiert aus einer unüberschaubar großen Datenmenge. Wie verhält sich diese neue mediale Konstellation zum klassischen Begriff der Reproduktion?

Trainings-Daten als (Re-)Produktion von Wirklichkeit

Big Data markiert die Ignoranz gegenüber einzelnen Bildaussagen, formalen oder gar künstlerischen Momenten von Bildern. Stattdessen stehen operationale Fragen im Vordergrund: Wie lässt sich das Abgebildete klassifizieren? Ist es ein Hund? Ist es eine Treppe? Ist es Paris, die Stadt, oder die Geschäftsfrau mit Nachnamen Hilton (vgl. Pereira/Moreschi 2020). Der komplexe fotografische Diskurs und seine zentralen Fragen nach Materialität, Zirkulation, Agency oder Situierung werden reduziert zugunsten der Operationalisierung der Bilder (Farocki 2004; Parikka 2023). Daraus folgt ein Strom trivialisierter Bilder, in dem das einzelne Bild seiner argumentativen Kraft beraubt ist. So stellen die Medientheoretiker Tomáš Dvořák and Jussi Parikka fest, dass Fotografie nicht länger als stabilisiertes Objekt gelten kann, welches Referenz erzeugt, stattdessen sei es „a different kind of image" geworden, beziehungsweise „a different kind of entity" (Dvořák/Parikka 2021, 12). Es folgt eine Verschiebung: weg von der Indexikalität und Zeug*innenschaft von Bildern hin zu Statistik, Wahrscheinlichkeitsverteilungen und Korrelationen. Es ist nicht länger die Kombination Auge/Kamera, welche das Sehen eines Bildes als Akt des Auswählens und Entscheidens und der bildlichen Sinngebung dominiert.

Aus dem Internet gescrapte Bilder sind nicht länger als Bild relevant, sondern allenfalls als zweidimensionale Kombination von Pixeln, wobei jeder einzelne Pixel eine gewisse Menge an abgestuften Farbwerten annimmt und mittels algorithmischer Verfahren operationalisiert wird. Das Indexikalische des Bildes wird hier zerlegt in atomisch kleinste Einheiten, in der Absicht, sie später erneut zu Pixelmengen zusammenzusetzen. Sie verorten sich gerade nicht im Diskurs des Fotografischen als Annäherung an die reproduk-

tive und indexikalische Wiedergabe von Wirklichkeit, sondern in deren brutaler Verneinung, im Feld des Operationalen und der Datenproduktion.

Ein Beispiel dafür ist die Bilddatensammlung LAION-5B. Sie besteht aus 5,85 Milliarden Bild-Text-Paaren, davon 2,3 Milliarden in englischer Sprache und 2,2 Milliarden Bild-Text-Paare in mehr als 100 weiteren Sprachen (Schuhmann u. a. 2022). Um diese Menge überhaupt denken zu können, hilft es, sich die Tausenden von Festplatten in einem oder mehreren Data Centern und deren ökologischen Fußabdruck durch Strom- und Wasserverbrauch vorzustellen (vgl. Holt/Vonderau 2015; Hogan 2024). Es sind Reihen und abermals Reihen, die aus sich wiederholenden Racks bestehen, in denen per Ethernet-Kabel verkettet die Serverrechner ihren Dienst tun.

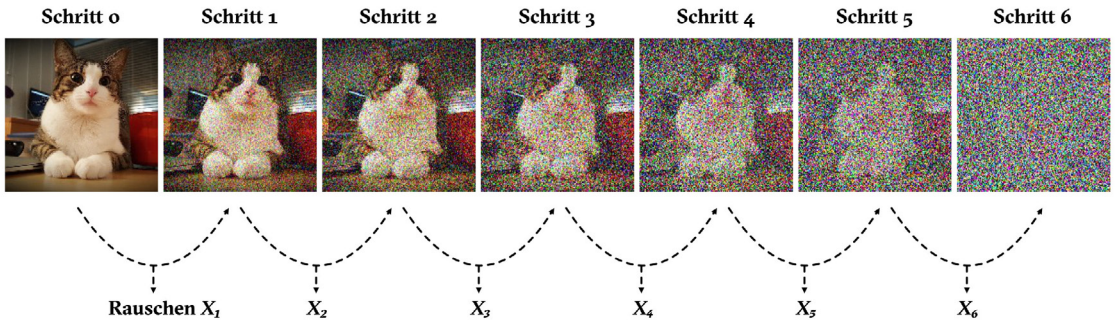
LAIONs Milliarden Bild-Text-Paare dienten u. a. zum Prägen von *OpenCLIP*, ein Bild-Text-Modell, welches bis dahin unbekannte Bilder prozessieren kann und die dazu wahrscheinlichste Bildbeschreibung ausgibt – eine Klassifikationsaufgabe. In planetarer Skalierung ist die Quelle das Internet, aus dem Informatiker*innen mittels der Software *Common Crawl* diese Datenmengen extrahieren. Dies können Wikipedia-Artikel, große Internet-Foren wie Reddit, die durch Google Books angelegten Textsammlungen oder öffentliche Bilddatenbanken wie Flickr sein (ebd.).

In seiner extraktivistischen Herkunft unterscheidet sich LAION leicht von früheren Datenerhebungsprojekten. Beim klassischen *ImageNet* wurden die Annotationen für jedes einzelne aus dem Internet extrahierte Bild von durch die Jobplattform Amazon Mechanical Turk vermittelten Clickworker*innen gegen schlechte Bezahlung vorgenommen (Li u. a. 2009; Denton u. a. 2021; Crawford/Luccioni 2024). Die Zuschreibungen, welche Pixelassemblagen (ich spreche hier absichtlich nicht mehr von Bildern) auf welche Zeichenketten (auch hier ist aus statistischer Sicht nicht mehr von Worten oder Sätzen zu sprechen) zutreffen, dienten dem zukünftigen Trainieren der Modelle. Als Datengrundlage für die Bild-Text-Paare bei LAION dienen die von den ursprünglichen User*innen kostenlos hergestellten Bildbeschreibungen, die im Internet mitveröffentlicht werden. Mit dem Verzicht auf bewusste Annotation und ausschließlichem Rückgriff auf das Scraping werden alle, die Inhalte ins Internet stellen, unfreiwillig zu unbezahlten Mitarbeiter*innen. Es entsteht eine unstrukturierte Datenmenge, deren wichtigste Eigenschaft es ist, sehr groß zu sein. Sie ist zusammengestellt entlang eines Paradigmas, welches Adrian Mackenzie und Anna Munster als *platform seeing* beschrieben haben: „images, not simply quantified, but labelled, formatted and made 'platform-ready'" (MacKenzie/Munster 2019, 5). Mit Andrew Dewdney and Katrina Sluis (2023) zielen Bilder in Sammlungen nicht länger auf das Einzelbild, sondern auf planetar zirkulierende „networked images" ab.

Transformer und Diffusion

Neben den Daten ist die zweite wesentliche Zutat die Architektur der gewichteten Netzwerke. Diese sind im Verlauf der letzten 15 Jahre komplexer geworden (zur Geschichte vgl. Offert 2022; zur Funktionsweise Somaini 2023). Die folgende, unvermeidlich komplexe Beschreibung erfolgt im Sinne einer Materialanalyse statistik-basierter generativer Verfahren zur Erstellung von Pixelassemblagen, welche die Künstlerin und Theoretikerin Hito Steyerl als „Mean Images" gekennzeichnet hat (Steyerl 2023).

Transformation bezeichnet wortwörtlich den Übergang einer Formation in eine andere, in diesem Fall die Umwandlung von Informationen (Bilder, Text, Ton) in Daten, die als mathematische Vektoren vorliegen. Transformer-Netzwerke wurden ab 2017 durch Google-Ingenieure entwickelt (Vaswani u. a. 2017). Sie beruhen darauf, Informationseinheiten, z. B. Wortfolgen, in kleinere Einheiten zu zerlegen. Ursprünglich auf das Feld der Sprache ausgerichtet, treffen Transformer Vorhersagen darüber, mit welcher Wahrscheinlichkeit ein bestimmter Textbaustein (ein ‚Token') auf einen vorhergehenden Textbaustein folgen werde. Je mehr Daten im Trainingsprozess in die Modelle eingeprägt werden, umso größer die Wahrscheinlichkeit, einen für Menschen sinnerzeugenden



| Abb. 1 | Vorwärtsprozess: Einem Ausgangsbild wird in mehreren Schritten Rauschen hinzugefügt. Das Rauschen X1...X6 in den verschiedenen Trainingsschritten dient dazu, ein U-Net für spätere Vorhersagen zu trainieren. Bild: Francis Hunger unter Verwendung von Kreis u. a. 2022, S. 16

Text zu generieren. Diese Steigerungslogik der Datenmenge gerät durch sinkende Grenzerträge an ihr Limit. Bei wachsendem Trainingsaufwand steigt der Ertrag nur wenig signifikant an. Zu den Grenzen des Verfahrens zählt auch die relativ limitierte Verfügbarkeit originärer (d. h. nicht-synthetischer) Daten, die sich inzwischen zunehmend mit synthetisch generierten Daten vermischen.

Damit generiert werden kann, müssen Transformermodelle zuerst einem Prozess des Einprägens unterzogen werden, der anthropomorphisierend als ‚Training' bezeichnet wird. Im Folgenden wird die Einprägephase und die Anwendungsphase für drei wesentliche Komponenten diskutiert: Erstens für ein CLIP Text-Bildmodell, in dem Pixelwahrscheinlichkeiten und Textwahrscheinlichkeiten in einem gemeinsamen Vektorraum eingeprägt werden. Zweitens ein Diffusionsmodell, das es erlaubt, synthetische Bilder aus der Umkehrung eines Bildrauschens zu generieren. Nach dem Einprägen gelten die Modelle als „pre-trained", und die Einprägungskomponente kann abgekoppelt werden. Drittens bleiben die komprimierten Vektorrepräsentationen der Ausgangsdaten übrig, die für das Generieren, das ‚Sampling', verwendet werden.

I. Für das Text-Bildmodell werden Worte in einem mehrdimensionalen Raum nach Ähnlichkeit verteilt, ein sogenanntes Embedding. Man kann sich diesen Raum dreidimensional mit x, y und z-Achse vorstellen, obwohl er in Wirklichkeit mehrdimensional ist und aus Tausenden Dimensionen bestehen kann. Aus der Verteilung von Worten untereinander wird für jedes Wort ein mathematischer Vektor abgeleitet, z. B. ‚Gemälde' (5, 9, 8). Andere Vektoren (x, y, z) liegen entweder in der Nähe von ‚Gemälde' (5, 10, 8) oder weiter entfernt (2, 2, 10). Je näher sie beieinander liegen, desto wahrscheinlicher ist es, dass sie statistisch einander zugehörig sind (‚Pinsel', ‚Farbe', ‚Rahmen'). Vektoren erlauben mathematische Operationen wie Addition und Subtraktion. Durch das statistisch berechenbare Verhältnis von Nähe und Richtung der Vektoren werden Bedeutungskorrelationen eingeschrieben. Eines der maßgeblichen Text-Bild-Modelle zur Vorhersage von Bildinhalten heißt *Contrastive Language-Image Pretraining* (CLIP) und wurde 2021 entwickelt (Radford u. a. 2021; vgl. auch Rodríguez-Ortega 2022; Offert 2023, 124).

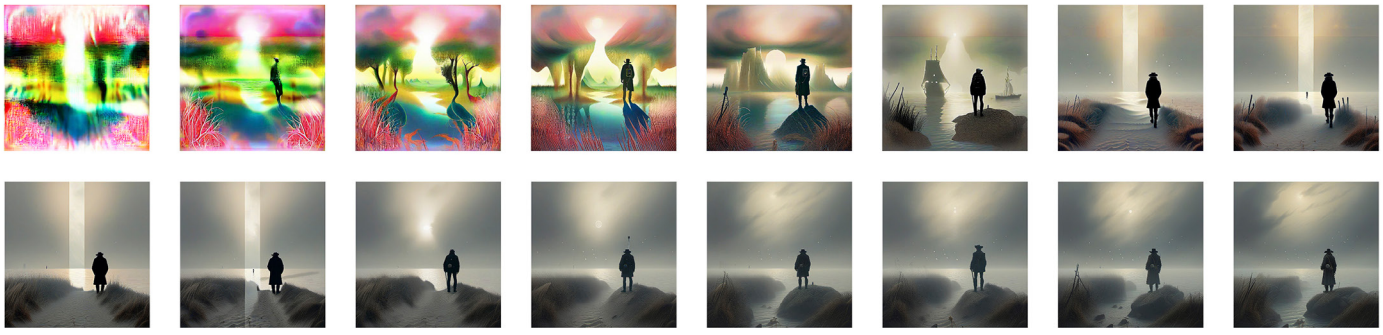
II. In generativen Diffusionsmodellen werden Daten im ‚Training' in zwei Richtungen eingeprägt, in einem Vorwärtsprozess und einem Rückwärtsprozess. Dieser Diffusions-Ansatz baute auf der Forschung eines Informatiker-Teams an der UC Berkeley (Ho/Jain/Abbeel 2020) und einer Kooperation zwischen der Stanford University und Google (Song u. a. 2021) auf und wurde zwei Jahre später durch ein Team am Lehrstuhl „KI für Computer Vision und Digital Humanities/die Künste" der LMU München als Open Source realisiert (Ommer u. a. 2022). Im Vorwärtsprozess wird in vielfachen Schritten zu einem Bild ein Gauß'sches Rauschen hinzugefügt, solange, bis es diffus ist und nur noch aus Rauschen besteht. | Abb. 1 | In jedem der circa 50 Schritte ist der Übergang von einem zum nächsten Bild bekannt und durch mathematische Parameter beschreibbar. Die aufeinander auf-

bauenden Schritte ergeben eine Markov-Kette des Verrauschens mit den Parametern ‚Mittelwert' des Rauschens, ‚Varianz' des Rauschens und ‚Schrittanzahl'. Die Ingenieure der Diffusionsverfahren fragten sich, ob sich das schrittweise Hinzufügen von Rauschen nicht umkehren lasse, um aus einem Rauschen schrittweise ein Bild zu generieren.

Für den Rückwärtsprozess stellten die Forscher*innen fest, dass sich der Diffusionsprozess nicht unmittelbar umkehren lässt, da er mathematisch nicht berechenbar ist. Statt exakter Berechnung wird basierend auf den im Vorwärtsprozess aufgezeichneten Rauschunterschieden ein Modell erstellt. Dieses gewichtete Netz (ein *Convolutional Neural Network* vom Typ ‚U-Net') wird so geprägt, dass es Vorhersagen über die beste Pixelkombination in den jeweiligen Rückwärtsschritten treffen kann. Als Maß wird die vektorielle Differenz zwischen dem vorhergesagten Rückwärtsübergang und dem aufgezeichneten Vorwärtsübergang mit den oben genannten Parametern verwendet. Bei bestimmten Parametern ist die Wahrscheinlichkeit einer bestimmen Pixelzusammenstellung in der Transformation am höchsten. Diese höchstwahrscheinlichen Transformationen werden für den generativen Prozess benötigt. | Abb. 2 | Dies wird für alle Bilder eines Datensatzes durchgeführt, um das U-Net zu prägen – im Fall von LAION-5B zum Beispiel für 5,85 Milliarden Bilder. Es entsteht ein vortrainiertes gewichtetes Transformernetzwerk, ein Modell.

III. Der generative Prozess, das Sampling, beginnt mit einem vollständigen Pixel-Rauschen. In circa 15 bis 20 iterativen Schritten wird Rauschen entfernt und zwar mittels des zuvor trainierten U-Net, welches die wahrscheinlichste Pixelmenge herausschält, die zunehmend als Bild erkennbar ist (Ho/Jain/Abbeel 2020, 2). Dies würde eine beliebige, für den Menschen als Bild lesbare Pixelassemblage erstellen. Doch ist das Ziel nicht ein beliebiges, sondern ein für Menschen spezifisches Bild. Als Guidance dient eine Texteingabe, ein Prompt. Dieser besteht aus einem oder mehreren Wörtern, die im oben skizzierten CLIP-Verfahren in Vektoren umgewandelt wurden. Mittels der Textvektoren wird eine vektorielle Richtung vorgegeben, in die das ebenfalls vektoriell encodierte Entrauschen (De-Noising) gelenkt wird. Für jeden Zustand eines Bildes wird also das wahrscheinlichste Muster zum Entrauschen berechnet, demnach es im Vektorraum in Richtung des Prompts zeigt, und zwar in so vielen Schritten, bis das Rauschen aus dem Bild eliminiert wurde. Erst im letzten Schritt werden die Vektoren wieder in eine Pixelzusammenstellung übersetzt (vgl. auch Chang u. a. 2023).

Zusammenfassend erscheint die generierte Pixelassemblage als Abstand zu einem Rauschen, indem der



| Abb. 2 | Schritt 1 bis 18 für den Prompt „Wanderer über dem Nebelmeer". Zwischen jedem Schritt wird mithilfe der Vorhersage des U-Nets Rauschen entfernt. Es ist nicht möglich, das letzte Bild direkt zu berechnen. In den ersten Schritten sind starke Veränderungen sichtbar, ab Schritt 7 sind sie graduell.
Stability AI Stable Diffusion SD-XL 1.0 (2023), Seed 0, 768 × 768 Pixel. Bild: Francis Hunger

| Abb. 3 | Straßenschlucht. Die uns als Perspektive erscheinenden Fluchtlinien sind keine Linien, die sich aus Anfangs- und Endpunkt ergeben, sondern statistisch berechnete Pixelkorrelationen. Prompt „Park Avenue and 34th Street NYC". Generiert mit Stability AI Stable Diffusion SD-XL 1.0 (2023), Seed 0, 20 Schritte, 768 × 768 Pixel. Bild: Francis Hunger

Diffusionsprozess die Wahrscheinlichkeitsverteilung der Trainingsdaten (wie zum Beispiel LAION-5B) modelliert.

Statistische Fläche

Synthetisch generierte Bilder sind keine Reproduktionen, auch wenn sie bestimmte optische Zusammenhänge ableiten. Einerseits werden einmal in die Transformernetzwerke eingeprägte optische Konfigurationen von einem Zufallsrauschen ausgehend rekonstruiert, bis sich eine Konfiguration ergibt, die den Trainingsdaten ähnelt, und zwar innerhalb jenes statistischen Raums, der durch bestimmte Wortkombinationen (Prompts) adressiert wird. Es sind „Anähnelungen". Andererseits sind es keine Rekonstruktionen im Sinne einer *Constructio*. Die dem Renaissance-Künstler Filippo Brunelleschi zugeschriebene „Entdeckung" der Zentralperspektive hat in westlichen Gesellschaften die *Constructio* als eine Formalisierungsleistung räumlicher Abstraktion formiert, die Menschen zum Erkennen realer räumlicher Verhältnisse dient. Derzeitige Transformernetzwerke (und gleiches gilt für alle derzeitigen ‚künstlichen', ‚neuronalen' Netzwerke) haben keinen Anlass für das Konzept ‚Constructio'. Sie verwenden, so Roland Meyer, „anders als etwa Games-Engines, Architektur-Renderings oder CGI-Effekte […] kein dreidimensionales Modell einer physischen Wirklichkeit, die nach optischen Gesetzen und den Regeln der Perspektive

| Abb. 4 | Zum Vergleich „Park Avenue and 34th Street. NYC". Digitale Fotografie, aufgenommen mit Apple iPhone 6 Plus, 2017. Foto: Carl Mikoy↗

berechenbar wäre“. Stattdessen synthetisierten sie „visuelle Texturen, Atmosphären und Anmutungen“ (Meyer 2022, 53). Der Grund ist simpel: Der Anlass dieser algorithmischen Verfahren liegt in mathematischen Optimierungen und nicht, wie beim menschlichen Körper, in der Notwendigkeit, sich im Raum orientieren zu müssen.

Transformernetzwerke, in deren Innerem Matrizen statistisch rechnen, erzeugen Flachheit. | Abb. 3 | und | Abb. 4 | Das, was vor unseren Augen ein Bild ergibt, ist aus mathematischer Sicht eine Anordnung von Pixeln, z. B. von 768 × 768 Höhe und Breite, also von insgesamt 589.824 Pixeln. Für jeden einzelnen dieser Pixel wird ein Farbwert berechnet, und zwar anhand der höchsten statistischen Wahrscheinlichkeit dafür, dass der Farbwert jenem Farbwert an der gleichen Position im Pixelraster entspricht, wie es bei den eingeprägten Trainingsdaten der Fall war. Werden mehrere Worte im Prompt miteinander kombiniert, zum Beispiel „van Gogh“ und „Astronaut“, so wird die Berechnung der statistischen Wahrscheinlichkeiten für die Pixelwerte miteinander korreliert. So entstehen die kombinatorischen Bildaussagen eines Van Gogh-Astronauten oder ähnlicher als surreal oder kitschig gelesener Bilder. | Abb. 5 | Auffällig im synthetischen Bild ist, dass ähnlich wie bei biometrischen Portraits die Augen in der Bildmitte fixiert sind, während die Mehrzahl der tatsächlichen Van Gogh’schen Selbstportraits eine seitlich gewendete Kopfhaltung zeigen. Die Textur des Pinselstriches wird stark übertrieben.

Die Betrachter*innen haben es mit einer optischen „Anähnelung“ zu tun, einer „Verflächigung“, die aus den mathematisch-statistischen Computingverfahren künstlicher ‚neuronaler‘ Netze folgt. Die Gemälde und Zeichnungen der Moderne waren auf die *Constructio* hin angelegt, die den Blick auf einen Fluchtpunkt lenkte. Perspektive setzt einen Sehpunkt voraus, motiviert durch einen menschlichen Körper und medialisiert durch Apparaturen. Diese Reproduktion von Raumverhältnissen entfällt. Die „symbolische Form der Perspektive“ (Panofsky 1927, 99) findet in Transformern nicht statt. Das allerdings hindert Menschen nicht daran, Raumverhältnisse in die flächigen Bilddaten hineinzulesen – eine menschlich-kognitive Rekonstruktion post factum. Meyer konstatiert: Es sind „keine Bilder der Welt, sondern Bilder aus Bildern“ (Meyer 2022, 53). Diesen gilt es nun nachzugehen, wozu die Metapher der Derivate, also der Ableitungen von Basiswerten, dienen soll.

Derivate

Derivate sind Finanzprodukte, die sich auf bestimmte Basiswerte beziehen, z. B. den Gold- oder Getreidepreis, aber auch Aktien, Schulden und Ähnliches. Derivate bestehen aus Einzelbestandteilen, die zerlegt und neu zusammengesetzt werden. Finanzderivate müssen zu einem bestimmten Zeitpunkt X in der Zukunft eingelöst werden, um spekulativ Gewinn oder Verlust zu erbringen. Auf den Zusammenhang von Derivaten und statistisch-generative Verfahren hat bereits der Physiker Dan McQuillan

| Abb. 5 | KI-Mansche. Prompt: „van gogh self-portrait in an astronaut suit in the style of van gogh“, generiert mit Stability AI Stable Diffusion SD-XL 1.0 (2023), Seed 0, 15 Schritte, 768 × 768 Pixel. Bild: Francis Hunger

hingewiesen: „Derivatives are a way of repackaging abstractions to provide a contractual foundation for financial speculation. Like AI, the derivatives market bets on correlations not on causations“ (McQuillan 2022, 55). Finanzderivate stellen die Vergleichbarkeit von Elementen über die Black-Scholes Differential-Gleichung her, welche neben Renditen und Zeit vor allem zukünftige Volatilität, das heißt die Stärke der Kurssprünge auf einen Zeitraum hin, modelliert. McQuillan argumentiert, dass in KI-Verfahren ebenfalls Vergleichbarkeit mittels differentialer Gleichungen erzeugt werde, die in Phasen als Reihung aufgerufen werden, z. B. im Verfahren der Backpropagation, dem Gradient-Descent-Algorithmus während des Trainings eines „KI“-Modells oder, wie oben diskutiert, den Rausch-Übergängen von Diffusionsmodellen.

Orit Halpern beschreibt in ihrer Diskussion der Black-Scholes Differential-Gleichung den Markt als „full of noise (as understood as unpredictable or not fully knowable signals)“, wobei die direkte Beziehung zwischen dem Derivat und dem Basiswert unbekannt sei. Allerdings können, so Halpern, die Varianzen der Aktienpreise im Zeitverlauf statistisch beobachtet werden und auch die Korrelation dieser Varianzen mit anderen Aktien (Halpern 2025). Ebenso wie die Volatilität der Aktienpreise im Verlauf von Zeit derivate Ableitungen auf Finanzmärkten erlaubt, ermöglichen die unterschiedlichen Rauschmuster eine Modellierung von Datenderivaten. Meine Analogie konzentriert sich daher auf den Aspekt der Ableitungen ausgehend von Basiswerten, konkret von aus dem Internet extrahierten Daten.

McQuillan verweist auf eine Logik der Fragmentierung, Finanzialisierung und Spekulation, die im Zuge der Neusynthetisierung von Bildern aus vorhandenen Trainingsdaten zum Tragen kommt. Voraussetzung ist die Dekomposition des vormals Ganzen, sodass Operationen auf kleinste Einheiten ausgeführt werden können, die frei von ihrem sozialen ‚Ballast‘ sind. Synthetische Bildproduktion eliminiert den komplexen Apparat der traditionellen Bildproduktion (Fotograf*innen, Maler*innen, Studios, Entwicklungs- und Vergrößerungstechniken, Distribution), der künstlerische Kreativität an Humankapital und damit laufende Kosten bindet. Wertschöpfung erfolgt hier entweder bei Clickworker*innen als Lohnarbeit, in deren Folge Arbeitszeit in den Wert der produzierten Ware wertschöpfend aufgeht. Alternativ wird der Wert der „Commons“ enteignet, indem die im Internet extrahierten Daten wie natürliche Ressourcen extrahiert werden, mit dem Ziel, sie zu Derivaten zusammenzusetzen (vgl. Hunger 2022, 217–222). Diese materialistische Interpretation zieht eine medientheoretische Wende in der Frage nach Original und Reproduktion nach sich.

Die Derivate sind also nicht ‚besser‘, ‚gleich gut‘ oder ‚schlechter‘ als das Original. Allenfalls sind sie spekulativ, wie etwa das obige Beispiel van Goghs im Astronautenanzug. Reproduktionen mittels traditioneller Vervielfältigungsverfahren sind hingegen an einen Index gebunden, an einen Ausgangspunkt, der im Feld des Symbolischen verankert ist, und erzeugen so Referenzen auf den Ursprungsgegenstand. Bild-Derivate profitieren vom Verächtlich-Machen des Originals: „with AI we have reproduction without original, more without one, a lie that doesn’t refer to any truth“ (Lorusso 2024). Synthetische „Bilder-aus-Bildern“ führen keinen Verweis auf ihre Herkunft mit sich, sie sind ein Abgesang auf die Mimesis (vgl. Steyerl 2023, 16). Beispielhaft kann das anhand ihrer Materialität diskutiert werden. So dokumentieren beim Holzschnitt die Materialität des Holzes, die handwerklichen und künstlerischen Fähigkeiten, die über einen längeren Zeitraum durch die Künstlerin erworben werden mussten, die Drucktechnik und das Reproduktionsmaterial (Holzart, Druckfarbe, Grammatur oder Färbung des Papiers) eine Kausalität zwischen Original und Reproduktion. Die Materialität determiniert in hohem Maße die Reproduktion. Bei derivativer Synthese ist dies nicht länger der Fall.

Die Umwandlung verschiedener Materialitäten in Vektoren, welche in den gewichteten Netzwerken statistischen Verfahren unterzogen werden, koppelt Materialität ab. Zusätzliche Verfahren zur Dimensionsreduzierung in neuronalen Netzen wie Softmax-Funktionen oder Principal Component Analysis

ignorieren Skalierungsinformationen und verstärken damit den Verlust von apparativer Indexikalität. Die Materialität des Outputs wird unabhängig von der Materialität des Inputs. Mit Diffusionsmodellen generierte Bilder sind Derivate basierend auf dem modellierten Entrauschen einer Pixelassemblage als Basiswert.

Fazit

Als synthetisch-generativ oder als derivative Assemblage gefasst, stellen generative Verfahren keine Reproduktionen im klassischen Sinne her. Generative Bilder sind Derivate auf den Durchschnitt der zugrundeliegenden Basiswerte, nicht aber auf ein konkretes Ausgangsereignis. Indexikalität existiert nur als Schatten vormaliger anderer Bilder, die in den Korrelationen des vektorialen Raums statistisch hinterlegt sind. Generative Bilder können daher allenfalls ikonographisch gelesen werden. Sie sind flach und perspektivlos. Während sich Finanzderivate auf mögliche zukünftige Ereignisse kaprizieren, kapitalisieren die Datenderivate der synthetischen Bildproduktion verausgabte Arbeit, die als naturalisiert erscheint.

Bibliographie

Altinoba 2020: Buket Altinoba, Das ‘Multiple’ im 19. Jahrhundert. Von Skulpturmaschinen, Techniktraktaten und Porträt-Miniaturbüsten, in: Dies. und Maria Männig (Hg.), Figuren der Replikation. kritische berichte 48/3, 2020, 67–80.↗

Barthes 1973: Roland Barthes, Le plaisir du texte, Paris 1973.

Benjamin [1935] 1991: Walter Benjamin, Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit [Erste Fassung], in: Gesammelte Schriften. Bd. 1: Abhandlungen, hg. v. Rolf Tiedemann und Hermann Schweppenhäuser, Frankfurt a. M. 1991, 431–469.

Bowker u. a. 2010: Geoffrey C. Bowker, Karen Baker, Florence Millerand und David Ribes, Toward Information Infrastructure Studies – Ways of Knowing in a Networked Environment, in: International Handbook of Internet Research, hg. v. Jeremy Hunsinger, Lisbeth Klastrup und Matthew Allen, Dordrecht 2010, 97–117.

Chang u. a. 2023: Ziyi Chang, George Alex Koulieris, Hyung Jin Chang und Hubert P. H. Shum, On the Design Fundamentals of Diffusion Models – A Survey, in: arXiv 2023.↗

Cox 2024: Geoff Cox, Photography at a Standstill, in: Media Theory 8/1, 2024, 297–318.↗

Crawford/Joler 2018: Kate Crawford und Vladan Joler, Anatomy of an AI System, Website 2018.↗

Crawford/Luccioni 2024: Kate Crawford und Alexandra Sasha Luccioni, The Nine Lives of ImageNet: A Sociotechnical Retrospective of a Foundation Dataset and the Limits of Automated Essentialism, in: Journal of Data-Centric Machine Learning Research 3, 2024, 1–18.↗

Denton u. a. 2021: Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart und Hilary Nicole, On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet, in: Big Data & Society 8/2, 2021.↗

Dewdney/Sluis 2023: Andrew Dewdney und Katrina Sluis (Hg.), The Networked Image in Post-Digital Culture, London/New York 2023.↗

Dvořák/Parikka 2021: Tomás Dvořák und Jussi Parikka (Hg.), Photography Off the Scale – Technologies and Theories of the Mass Image (Technicities), Edinburgh 2021.

Farocki 2004: Harun Farocki, Phantom Images, in: Public 29, January 2004, 12–22.↗

Halpern 2025: Orit Halpern, Financializing Intelligence – On the Integration of Machines and Markets, in: carrier-bag.net. Revised reprint of 2023, 2025.↗

Ho/Jain/Abbeel 2020: Jonathan Ho, Ajay Jain und Pieter Abbeel, Denoising Diffusion Probabilistic Models, in: arXiv 2020.↗

Hogan 2024: Mel Hogan, Artificial Intelligence Is a Hot Mess, in: Training the Archive, hg. v. Inke Arns, Eva Birkenstock, Dominik Bönisch und Francis Hunger, Köln 2024, 33–55.

Holt/Vonderau 2015: Jennifer Holt und Patrick Vonderau, „Where the Internet Lives“ – Data Centers as Cloud Infrastructure, in: Signal Traffic – Critical Studies of Media Infrastructures, hg. v. Lisa Parks und Nicole Starosielski, Urbana 2015, 71–93.↗

Hunger 2022: Francis Hunger, Data Workers of All Countries, End It!, in: Hamburg Maschine – Digitalität, Kunst und urbane Öffentlichkeiten, hg. v. Isabella Kohlhuber und Oliver Leistert, Hamburg 2022, 98–139.

Hunger 2023: Francis Hunger, Unhype Artificial Intelligence! A Proposal to Replace the Deceiving Terminology, in: Training the Archive – Working Paper 6, Aachen/Dortmund, April 2023.↗

Koebler 2025: Jason Koebler, AI Slop Is a Brute Force Attack on the Algorithms That Control Reality, in: Online Magazin 404 Media, 17. März 2025.↗

Kreis u. a. 2022: Karsten Kreis, Ruiqi Gao und Arash Vahdat, Denoising Diffusion-Based Generative Modeling: Foundations and Applications. Presented at the Computer Vision and Pattern Recognition Conference (CVPR), New Orleans 2022.↗

Li u. a. 2009: Fei-Fei Li, Jia Deng, Wei Dong, Richard Socher, Li-Jia Li und Kai Li, ImageNet: A Large-Scale Hierarchical Image Database, in: IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL 2009, 248–255.↗

Lorusso 2024: Silvio Lorusso, Deepdreaming Willy Wonka: AI Weird as the New Kitsch, Website 2024.↗

MacKenzie/Munster 2019: Adrian MacKenzie und Anna Munster, Platform Seeing – Image Ensembles and Their Invisualities, in: Theory, Culture & Society 36/5, 2019, 3–22.↗

McQuillan 2022: Dan McQuillan, Resisting AI: An Anti-Fascist Approach to Artificial Intelligence, Bristol 2022.

Meyer 2022: Roland Meyer, Im Bildraum von Big Data. Unwahrscheinliche und unvorhergesehene Suchkommandos: Über Dall-E 2, in: Cargo 55, 2022, 50–53.

Meyer 2025: Roland Meyer, Echte Emotionen. Generative KI und rechte Weltbilder, in: Geschichte der Gegenwart 2, Februar 2025.↗

Offert 2022: Fabian Offert, Ten Years of Image Synthesis. Blogbeitrag auf zentralwerkstatt.org vom 10. November 2022.↗

Offert 2023: Fabian Offert, On the Concept of History (in Foundation Models), in: IMAGE. Zeitschrift für Interdisziplinäre Bildwissenschaft 19, 2023, 121–134.↗

Ommer u. a. 2022: Björn Ommer, Robin Rombach, Andreas Blattmann, Dominik Lorenz und Patrick Esser, High-Resolution Image Synthesis with Latent Diffusion Models, in: arXiv 2022.↗

Panofsky 1927: Erwin Panofsky, Die Perspektive als symbolische Form, in: Aufsätze zu Grundfragen der Kunstwissenschaft, hg. v. Eugen Verheyen und Hariolf Oberer, Berlin 1927, 99–167.

Parikka 2023: Jussi Parikka, Operational Images: From the Visual to the Invisual, Minneapolis, MI 2023.

Pasquinelli 2023: Matteo Pasquinelli, The Eye of the Master: A Social History of Artificial Intelligence, London/New York 2023.

Pereira/Moreschi 2020: Gabriel Pereira und Bruno Moreschi, Artificial Intelligence and Institutional Critique 20 – Unexpected Ways of Seeing with Computer Vision, in: AI & SOCIETY, September 2020.↗

Radford u. a. 2021: Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, Learning Transferable Visual Models From Natural Language Supervision, in: arXiv 2021.↗

Rodríguez-Ortega 2022: Nuria Rodríguez-Ortega, Techno-Concepts for the Cultural Field: N-Dimensional Space and Its Conceptual Constellation, in: Multimodal Technologies and Interaction 6/11, 2022, 96.↗

Schuhmann u. a. 2022: Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, Jenia Jitsev, LAION-5B – An Open Large-Scale Dataset for Training next Generation Image-Text Models, in: arXiv 2022.↗

Siegel 2020: Steffen Siegel, Nicéphore Niépce und die Idee der fotografischen Replikation, in: Buket Altinoba und Maria Männig (Hg.), Figuren der Replikation. kritische berichte 48/3, 2020, 95–106.

Somaini 2023: Antonio Somaini, Algorithmic Images: Artificial Intelligence and Visual Culture, in: Grey Room 93, Oktober 2023, 74–115.↗

Song u. a. 2021: Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon und Ben Poole, Score-Based Generative Modeling through Stochastic Differential Equations, in: arXiv 2021.↗

Steyerl 2023: Hito Steyerl, Mean Images, in: New Left Review 140/141, Juni 2023.↗

Ullrich 2025: Wolfgang Ullrich, KI-Bilder: Wenn Bilder nur noch Bullshit sind, in: Die ZEIT Online vom 7. Januar 2025.↗

Vaswani u. a. 2017: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser und Illia Polosukhin, Attention Is All You Need, in: arXiv 2017.↗