Zeitschrift für Medienwissenschaft, Jg. 17, Heft 33 (2/2025), https://doi.org/10.14361/zfmw-2025-170212. Published by transcript Verlag.

FRANZISKA BARTH

RESIDUAL IMAGES

Im Frühjahr 2023 lässt sich die Medienkünstlerin Hito Steyerl auf ein folgenreiches Experiment ein. Mithilfe des von Mathew Dryhurst und Holly Herndon entwickelten Website-Tools *Have I Been Trained?*[1] hatte sie herausgefunden, dass die LAION-5B-Datenbank, die maßgeblich zur Schulung von generativen Bildmodellen wie etwa Stable Diffusion verwendet wird, sowohl Bilder ihrer künstlerischen Arbeiten als auch ihrer Persona enthielt. Daraufhin forderte sie das Modell mit dem Prompt «Image of Hito Steyerl» dazu auf, ein Bild von sich zu generieren (vgl. Abb. 1). Mit dem Ergebnis war sie denkbar unzufrieden. Jeglicher Versuch, das von Stable Diffusion generierte Bild zu beschreiben, muss sich mit hoher Wahrscheinlichkeit auf ein Vokabular berufen, das eng mit der Beschreibung von Karikaturen assoziiert ist. So erscheinen die ohnehin überzeichneten Gesichtszüge der hier dargestellten Steyerl durch die dramatische Lichtsetzung deutlich zerfurchter, das Haar merklich ungeordnet. Statt einer adäquaten Repräsentation sieht sich die Autorin mit einer verzerrten Darstellung, «in a state of frozen age range, produced by internal, unknown processes, spuriously related to the training data», konfrontiert.[2]

In ihrer anschließenden essayistischen Reflexion über die grundsätzliche Befremdlichkeit generierter Bilder, die sie im Angesicht ihres missglückten Porträts als «averaged versions of mass online booty» bezeichnet,[3] entwickelt sie den Begriff der «mean images», der ‹gemeinen Bilder›. Mit dem englischen Wort *mean* spielt sie dabei vor allem auf dessen inhärente Doppellogik an. Einerseits bezieht es sich auf das ‹Gemeine› im Sinne des ‹Mittelmäßigen› und ‹Durchschnittlichen› sowie das unabdingbare ‹Mittel zum Zweck› (*means to an end*) und verweist damit auf die Tatsache, dass diese Bilder mit Blick auf ihre Produktionsbedingungen das Ergebnis eines stochastisch induzierten Mittelwerts sind; andererseits steht *mean* für das ‹Bösartige›, ‹Niederträchtige› und ‹Schäbige› – all jene Eigenschaften, die sich nach Steyerls Ansicht auf der sicht-

1 *Have I Been Trained* ist ein öffentlich zugängliches Online-Tool, das es Nutzer*innen ermöglicht, herauszufinden, ob bestimmte Daten oder Bilder von generativen Modellen zu Trainingszwecken verwendet wurden. Diese Daten lassen sich bei Bedarf durch eine Opt-out-Funktion aus dem Trainingsdatensatz ausschließen. Vgl. *haveibeentrained.com* (31.3.2025).

2 Hito Steyerl: Mean Images, in: *New Left Review*, Nr. 140/141, 2023, 82–97, hier 84.

3 Ebd., 82.

Abb. 1 **Mit Stable Diffusion generiertes Bild zum Prompt «Image of Hito Steyerl»**

baren Oberfläche dieser Bilder niederschlagen, als «an approximation of how society, through a filter of average internet garbage, sees me».[4]

Das deformierte Selbstbild von Stable Diffusion wird für sie zu einer dekodierbaren und somit lesbaren Oberfläche mit «dokumentarischem Ausdruck»,[5] in der sich die ausbeuterische Logik der Produktionsbedingungen dieser Bilder, soziale Ungleichheiten und Diskriminierung visuell Ausdruck verschaffen. Der Begriff der *mean images* verfängt dabei zweifelsohne nicht nur aufgrund der rhetorischen Gewandtheit der Autorin, sondern weil gerade auch das Bild, das sie zu diesem Begriffskonzept verleitet, auf ganz eigenwillige Weise für sich zu sprechen scheint – *to show rather than to tell*. In jenem Moment nämlich, da die bildliche Abweichung zur realen Persona Steyerls durch den karikaturhaften, aber dennoch fotorealistischen Duktus zum (ungewollten) Träger einer politischen Botschaft wird.

Was durch die vermeintliche Selbstverständlichkeit ihres visuellen Arguments jedoch in den Hintergrund rückt, ist die Frage nach seiner Folgerichtigkeit. Denn das grundsätzlich ‹neue› Potenzial dieser Bildgeneratoren liegt ja gerade nicht in der *Reproduktion* des Identischen, sondern in der *Produktion* des Generischen. Die Niederträchtigkeit, die Steyerl diesen Bildern zuschreibt, gründet deshalb möglicherweise nicht in der Abbildung selbst, sondern entsteht aus der Dissoziation zu einer spezifischen Erwartungshaltung gegenüber dem, was ein «Image of Hito Steyerl» leisten soll: eine repräsentative Ähnlichkeitsbeziehung.

Der implizite Anspruch an diese Bilder, so möchte ich argumentieren, ist das Erbe eines fotografischen Realismus, der sich jedoch in veränderter Form, nämlich auf der Basis einer quantitativen Logik in diese Bilder einschreibt. Der seltsame Beigeschmack jener Bilder stünde somit weniger exemplarisch für eine

4 Ebd., 84.
5 Ebd., Übers. FB.

allgemeine ‹Niederträchtigkeit› generierter Bildwelten als vielmehr für ein Rumoren in der Subduktionszone widersprüchlicher Realismus-Konzepte.

Für eine Annäherung an diese Problematik schlage ich deshalb anschließend an Steyerls Begriff der *mean images* den Begriff der *residual images* vor. Das Residuale steht dabei im Sinne des Restlichen oder Resthaften für zweierlei Funktionen: Zum einen ermöglicht der Begriff, dem evozierten Abfallcharakter dieser Bilder eine ambivalentere Lesart jenseits binärer Kategorien zu unterziehen, um letztlich in dieser Eigenschaft auch ein analytisches Potenzial zu verorten. Andererseits ermöglicht er, die Bilder im Kontext der spezifischen Ökologie dieser Modelle als zentrifugale Reste zu adressieren, deren Gravitationszentrum sich nach wie vor um fotografische Paradigmen des (Aus-)Sortierens formiert.

I Fotorealismus von unten: die Log(ist)ik der Halde

Zu Beginn noch einmal eine kurze Definition: Text-zu-Bild-Generatoren wie Stable Diffusion, DALL-E oder Midjourney sind systemische Plattformen, die dazu in der Lage sind, innerhalb von Sekunden fotorealistische Bilder auf Basis sprachlicher Eingaben, sogenannten Prompts, zu erstellen. Technologisch basieren diese Systeme auf den Prinzipien des maschinellen Lernens und nutzen dafür Architekturen wie Generative Adversarial Networks (GANs) oder latente Diffusionsmodelle. Was diesen Modellen ermöglicht, auf der Basis einer sprachlichen Kondition – dem *prompting* – Bilder zu generieren, ist dabei zunächst ein von OpenAI entwickeltes Modell namens CLIP (Contrastive Language-Image Pre-training), das Sprache und Bilder miteinander verknüpft.[6]

Die materielle Existenzgrundlage von Bildgeneratoren (wie auch aller anderen generativen Modelle) basiert deshalb auf der exzessiven Ansammlung von (überwiegend) fotografischen Bildern – einer Art Daten-Hoarding. Was demnach dazu verleitet, von den Trainingsdaten dieser Modelle als *internet garbage* zu sprechen, ist nicht notwendigerweise die tatsächliche Qualität bzw. der Wert einzelner Bilder, sondern vielmehr die Tatsache, *wie* diese Bilder diskursiv in Erscheinung treten. Einen Anlaufpunkt bildet dabei bereits die Benennung der Datenkonvolute an sich. So erweist sich beispielsweise die Bezeichnung LAION 5B trotz der phonetischen Anspielung auf den Löwen als Sinnbild von Herrschaft und Ordnung als ein anästhetischer Rekurs, der vielmehr auf die Absenz von Regentschaft verweist: Denn LAION 5B ist schlichtweg das Akronym für Large-scale Artificial Intelligence Open Network mit fünf Milliarden Datensätzen.[7] Der tatsächliche *wildlife encounter* besteht also allein im Moment der schieren Überschreitung des menschlichen Maßstabs, in der Unbändigkeit eines Archivs, das sich gerade dadurch auch als Archiv infrage stellt. Dies hängt nicht zuletzt mit der Art und Weise der Sammlungsakquise zusammen: Milliarden digitaler Fotografien werden durch sogenannte Webcrawler oder Spiders gesammelt – Scraping-Programme,

6 Vgl. Alec Radford u. a.: CLIP: Connecting text and images, *OpenAI*, 5.1.2021, openai.com/index/clip (31.3.2025).

7 LAION 5B ist ein Projekt der Nonprofit-Organisation LAION mit insgesamt 5,85 Milliarden CLIP-gefilterten Text-Bild-Paaren, vgl. Romain Beaumont: LAION-5B: A New Era of Open Large-Scale Multi-Model Datasets, *LAION.ai*, 31.3.2022, laion.ai/blog/laion-5b/ (31.3.2025).

die das frei zugängliche World Wide Web bis in die hintersten Winkel selbstständig durchstöbern und indexieren –, um Ansammlungen oder bloße ‹Haufen› von Bildern aufzuschichten, die keinem spezifischen Selektionsprozess und keiner weiteren Qualifizierung jenseits der Formel ‹X ist ein Bild von Y› unterliegen.

Der faulige Beigeschmack dieser Bilder geht dementsprechend nicht auf eine allgemeine, intrinsische Qualität der Trainingsbilder zurück, die sich eins zu eins in ihnen Ausdruck verleiht, sondern emergiert einerseits aus der Art und Weise ihrer Wertschöpfung und andererseits aus einem damit verbundenen Aspekt der Unordnung. In der Regellosigkeit, im wahllosen Neben- und Übereinander entstehen prekäre Anordnungen, in der die zuvor sorgfältig getrennten semantischen Einheiten aus Cat Content, Werbefotografie, Kunst, Privatem, Kinderpornografie und Gewaltdarstellungen in ihrer bedrohlichen Unförmigkeit zu einer amorphen Masse, einer Art ‹Resterampe› unserer kollektiven Bildwelten denaturieren. Und obwohl die explizit verletzenden Inhalte rein quantitativ betrachtet eine Randerscheinung darstellen, scheinen sie gerade durch ihre Ortlosigkeit im *latent space*[8] der Unkontrollierbarkeit stets die Gefahr einer ökologischen Verunreinigung und somit einer Kontamination des gesamten Milieus zu beschwören. Das dadurch aufkommende Moment des grundlegenden Verdachts, das spätestens seit der Theoretisierung der digitalen Fotografie beständig in den Vordergrund rückt,[9] beschränkt sich dementsprechend längst nicht mehr nur auf bildimmanente fotorealistische Kategoriensysteme wie *real* und *fake*, sondern greift spätestens mit den generativen Bildern tief in die Schichten gesellschaftlicher Deutungssysteme, Wertvorstellungen und moralischer Ordnungen ein.

In der Funktion der Bildgeneratoren, potenziell Unvereinbares miteinander in Verbindung zu bringen, liegt also zweifelsohne eine datenökologische *toxicity*, die an dieser Stelle keinesfalls verharmlost werden soll. Dennoch lässt sich die *messiness* dieser Bilder nicht essenzialisierend auf die unmittelbare Qualität der Trainingsdaten zurückführen. Sie ist vielmehr die Konsequenz einer komplexen Bildlogistik, in der Bilder sowie die Arbeit an Bildern entwertet und ganz buchstäblich ‹wie Abfall› behandelt werden und dementsprechend auch ‹wie Abfall› zueinander in Beziehung treten können. Im Zeitalter des Anthropozäns liegt es jedoch auf der Hand, dass ein Nachdenken über diese materielle Zudringlichkeit nicht einfach in einer bloßen Forderung nach Ausschluss enden kann.

In seiner *rubbish theory* entwickelt der Mathematiker und Anthropologe Michael Thompson Ende der 1970er Jahre eine eigenständige Theorie aus dem Geiste des Mülls, die den Müll vor allem als statische Wertekategorie infrage stellt und ihn stattdessen als zentrales und vor allem zyklisch strukturiertes *liminal object* des Ambivalenten und Widersprüchlichen ins Feld führt.[10] Der konzeptuelle Fluchtpunkt dieser zyklischen Logik ist dabei zunächst in der ökonomischen Sphäre zu verorten, insofern er sich in der kulturellen Aufwertung

8 Als *latent space* wird ein mathematisch modellierter, meist niedrigdimensionaler Raum bezeichnet, in dem Datenmerkmale als kompakte Repräsentationen (latente Variablen) codiert werden. In Deep-Learning-Architekturen wie generativen Modellen dient der *latent space* dazu, komplexe Eingabedaten (z. B. Bilder oder Texte) auf ein reduziertes, strukturierbares Format abzubilden, das semantische Eigenschaften bewahrt und Manipulationen erlaubt. Vgl. Jaroslaw Drapala: What Is a Latent Space? A Concise explanation for the general reader, *Medium*, 8.5.2024, medium.com/data-science/what-is-a-latent-space-065eb8e3f859 (17.06.2025).

9 Vgl. Roland Meyer (Hg.): *Bilder unter Verdacht. Praktiken der Bildforensik*, Berlin 2024.

10 Vgl. Michael Thompson: *Rubbish Theory: The Creation and Destruction of Value*, London 2017 [1979], 4.

des *rubbish* durch Wiederverwertung manifestiert. Weniger besprochen sind indes Thompsons unmittelbar daran anschließende Bestrebungen, aus dem mathematisch-logischen Nachdenken über dinghafte Kategoriensysteme eine umfassende Gesellschaftstheorie abzuleiten, die ihn unter der sprechenden Kapitelüberschrift «monster conservation» dazu führt, die konkreten sozialen Praktiken und Diskurse ins Auge zu fassen, die sich um ‹monströse›, kategorial herausfordernde Dinge formieren. Nicht nur, so Thompson, offenbare sich in ihnen das transformatorische Potenzial jener Dinge, sondern sie bildeten auch einen Gradmesser für den jeweiligen Umgang von Gesellschaften mit Unordnung und Ambiguitäten.[11]

Ein weitaus ‹vitaleres› Verständnis von *messiness* und der grundsätzlichen Agency von abfälligen Prozessen wie Ablagerung, Zersetzung, Fäulnis oder Fermentation findet sich vor allem in den neumaterialistischen Ansätzen der Wissenschaftshistorikerin Donna Haraway. In *Staying with the Trouble* lädt sie dazu ein, das als ‹Müll› behandelte Abgestorbene, Aussortierte, Verdrängte grundsätzlich als Kompost und das Kompostieren stets als *lively art* – einen schöpferischen Akt und zutiefst chaotischen und relationalen Prozess – zu denken.[12] Die Text-zu-Bild-Generatoren mit Haraway als eine Form der ‹Kompostierungsmaschine› zu lesen, mag dabei auf den ersten Blick wie ein verfehlter Animismus wirken, doch diese Lesweise stützt sich letztlich auf das zentrale systemtheoretische Argument von M. Beth Dempster, auf das sie sich bezieht. Demnach stünde Sympoiesis für «collectively producing systems that do not have self-defined spatial or temporal boundaries. Information and control are distributed among components. The systems are evolutionary and have the potential for surprising change.»[13]

Generative Technologien wären demnach als zutiefst verstrickt in unsere Lebenswelt zu verstehen, in die sich das Modrige, Toxische und Unkontrollierbare von vornherein als eine Qualität des Abgelegten und Abgelebten einschreibt. Liegt also in der Auseinandersetzung mit diesen Bildern nicht auch ein Potenzial, mit dem sich Fragen jenseits von Kategorien wie Wert und Unwert formulieren lassen?

Was die generativen Modelle im Unterschied zu bisherigen Bildtechnologien für eine solche Betrachtungsweise prädestiniert, ist die Tatsache, dass sich die konkreten ‹Reaktionsprozesse›, die zur Formgenese führen, unserer Wahrnehmung entziehen. Während wir beispielsweise das ‹Bildprogramm› der traditionellen oder digitalen Fotografie in seiner sequenziellen Logik zumindest theoretisch nachvollziehen können, macht sich das generative Modell im wahrsten Sinne ein ‹eigenes Bild› der Fotografie. Die Regeln zur Formgenese sind dem Bildprozess dementsprechend nicht mehr äußerlich, sondern sie sind lediglich implizit im generierten Ergebnis vorhanden. Diese beobachtbare Verschiebung, die mit der Etablierung Künstlicher Neuronaler Netze (KNNs) im Bereich generativer Technologien einhergeht, bezeichnet Hannes Bajohr als «konnektionistische[s] Paradigma»:

11 Vgl. ebd., 138–157.
12 Vgl. Donna J. Haraway: *Staying with the Trouble: Making Kin in the Chthulucene*, Durham 2016.
13 Ebd., 61.

> Es stehen also in KNNs keine Herstellungsweisen, sondern Daten am Anfang, und aus ihnen wird erst durch einen iterativen Lernprozess das Modell gebildet; dieses Modell wiederum ist kein Algorithmus, sondern beschreibt lediglich die Verbindungsstärken zwischen den ‹Neuronen› in einem sogenannten Gewichtungsmodell.[14]

An anderer Stelle formuliert Bajohr dazu eine geradezu bildhafte Beschreibung, die womöglich nicht ganz zufällig jene Metaphorik eines ‹Rumorens› aus dem *uncommon ground* aufnimmt, auf die bereits Bezug genommen wurde: «Es [das Modell] ‹denkt sich› also, könnte man sagen, von Kanten über einfache Formen bis zu komplexen Objekten ‹hinauf›».[15]

Auch die konkrete Funktionsweise von Diffusionsmodellen wie Stable Diffusion reiht sich in diese Metaphorik ein. Hier entstehen die Bilder durch *denoising*, einen iterativen Prozess des Entrauschens. Zuvor wurde das Modell in einem *forward process* darauf trainiert zu ‹schätzen›, wie sich beispielsweise das Bild einer Banane durch das schrittweise Hinzufügen von Rauschen in einen vollständigen Rauschzustand wandelt, um daraus letztlich eine Wahrscheinlichkeit für die Rückoperation – vom verrauschten Bild zur gegebenen Kondition (in diesem Fall ein Bild, das zum Text «Bild einer Banane» passt) – abzuleiten.[16] Bildlich gesprochen lernen diese Modelle also vor allem, das Bild einer Banane ex negativo zu generieren – indem sie vom Zustand des reinen Rauschens all das abziehen, was statistisch gesehen *kein* Bild einer Banane ist. Mathematisch gesehen scheint es folglich effizienter zu sein, sich dem Bild über die Menge des vom Bild Ausgeschlossenen, von seinem Rest her zu nähern.

Was sich hier also hinsichtlich bestehender Paradigmen des Fotografischen verschiebt, ist die Tatsache, dass es für die Herstellung dieser fotorealistischen Bilder kein explizites ‹Bildäußeres› mehr bedarf und das Bild einer Banane zum selbstbezüglichen Kompositum jenes Inputs wird, der – heutzutage oftmals selbst durch ein Programm – als Bild einer Banane gelabelt wurde. Konsequenterweise muss in dieser Umkehrung, d. h. im Rückschluss vom Ergebnis auf die zugrunde liegenden Regeln, eine *andere* Bildlichkeit bzw. eine andere Form des Fotorealismus aus den rauschenden Tiefen der Datenhalden emportauchen als jene, die unseren bisherigen Sehgewohnheiten entspricht. Was sich hier formiert, ist ein Fotorealismus aus dem Geist seiner Residuen.

II Fotorealismus von oben: fotografische Residuen

Es gibt folglich gute Gründe, generierte Bilder als eine eigenständige Bildform zu begreifen, die sich grundsätzlich von der Fotografie unterscheidet. Dass man allerdings so weit geht, sie schlichtweg aus dem Gewebe der Geschichte und Theorie der Fotografie herauszuschneiden,[17] erfordert eine sehr verengte Sichtweise auf diese Modelle, die sich maßgeblich für technologische Infrastrukturen interessiert und weniger für die symbolischen Formen, mit denen die Modelle operieren. Amanda Wasielewskis entschlossenes Plädoyer für eine

14 Hannes Bajohr: Künstliche Intelligenz und digitale Literatur. Theorie und Praxis konnektionistischen Schreibens, in: ders., Annette Gilbert: *Digitale Literatur* II, München 2021, 174–185, hier 177.

15 Hannes Bajohr: *Schreibenlassen. Texte zur Literatur im Digitalen*, Berlin 2022, 166.

16 Vgl. Jonathan Ho, Ajay Jain, Pieter Abbeel: Denoising Diffusion Probalistic Models, *arxiv.org*, 19.6.2020, überarb. 16.12.2020, doi.org/10.48550/arXiv.2006.11239 (31.3.2025).

17 Seit der öffentlichkeitswirksamen Ablehnung des Sony World Photo Award im Jahr 2023 für das von ihm eingereichte KI-generierte Bild spricht sich der Philosoph und Künstler Boris Eldagsen entschieden gegen eine Vermischung der Kategorien aus, da generierte Bilder keine Fotografien seien. Vgl. Victor Sattler: Fotograf Boris Eldagsen über Künstliche Intelligenz: «Für mich ist KI eine Befreiung», in: *Monopol. Magazin für Kunst und Leben*, 23.11.2023, monopol-magazin.de/interview-boris-eldagsen (31.3.2025).

Einbindung der generativen Bildproduktion in fotografietheoretische Diskurse hat die diesbezügliche Debatte maßgeblich befördert.[18] Dabei wird sie nicht müde zu betonen, dass es nicht darum gehen kann, generierte Bilder eindeutig einer Bilderklasse zuzuordnen, sondern vielmehr darum, sich entlang der taxonomischen Grenzen auf Spurensuche zu begeben und gegebenenfalls auch das Verständnis des Fotografischen selbst einer Aktualisierung zu unterziehen: «Debates on taxonomy […] can also be an exercise in shifting expectations, stretching preconceptions and moving between the general and the specific in the implicit taxonomies shared within communities.»[19]

Ein wichtiges Stichwort ist hier sicherlich der Begriff der Präkonzeption. Denn es darf in der Betrachtung der skizzierten Modellökologie nicht vergessen werden, dass zu Beginn eines jeden Bildprozesses zwar eine ‹leere Zeile› steht, diese aber kein *blank space* im eigentlichen Sinne ist. Der semantische ‹Freiraum›, welcher den User*innen zur Verfügung steht, ist bereits vielfach codiert und es lässt sich durchaus behaupten, dass es vor allem fotografische Konditionen sind, die diesen Raum strukturieren – allein deshalb, weil der Fotorealismus einen offensichtlichen Qualitätsmaßstab für die Selbstbeschreibung dieser Modelle darstellt. So wurde beispielsweise der Launch des Modells SDXL 1.0 von Stability AI im Juli 2023 damit beworben, dass es sich um das derzeit beste Modell zur Erstellung fotorealistischer Bilder handele, wobei hier vor allem auf die vermeintliche ‹Reinheit› oder ‹Echtheit› dieses Realismus verwiesen wird: «without having any particular ‹feel› imparted by the model».[20]

Diese Präkonzeption, die sich in der Kammerszene des Promptings Ausdruck verschafft, muss jedoch im Kontext einer wesentlich weitreichenderen Entwicklung im Feld der Computer Vision betrachtet werden. Genau genommen geht es um die dort nach wie vor ungebrochen vorherrschende Bewertung fotografischer Bilder als *natural* oder *realistic images*. Zur Illustrierung seien hier zwei Beispiele genannt: So verwendet der Informatiker Daniel Ruderman 1994 in seinem damals wegweisenden Paper «The Statistics of Natural Images» durchweg einen Begriff der *natural images*, der zwar vage definiert, jedoch an keiner Stelle technologisch determiniert wird: «[Natural images] are far from random: images constructed randomly on a computer practically never contain a naturalistic scene – or even a tree. Natural images are thus very rare among the huge space of all possible images.»[21] Erst durch das gegebene Bildbeispiel «Image from the woods: rocks in a stream with background foliage»[22] lässt sich *induktiv* schließen, dass er sich auf die konkrete Bildform fotografischer Naturdarstellungen bezieht. Auch das vielzitierte Paper «ImageNet: A large-scale hierarchical image database» aus dem Jahr 2009 kommt bezüglich seines Bildbegriffs gänzlich ohne das Wort *photography* aus, obwohl es sich bei allen dargestellten Beispielbildern mit großer Wahrscheinlichkeit um (digitale) Fotografien handelt.[23] Diese induktive Gleichsetzung von *(natural) images* und Fotografie in den Naturwissenschaften hat eine lange zurückreichende Geschichte, auf die an dieser Stelle nur hingewiesen werden kann.

18 Vgl. Amanda Wasielewski: Unnatural Images: On AI-Generated Photographs, in: *Critical Inquiry*, Bd. 51, Nr. 1, 2024, 1–29, doi.org/10.1086/731729; vgl. die in der *Critical Inquiry*, Bd. 51, Nr. 2, 2025 erschienenen Reaktionen auf Wasielewskis Beitrag von Brooke Belisle, Ina Blom und Matthew Fuller, Marc Downie, Avery Slater.

19 Amanda Wasielewski: Critical Response V: AI-Generated Images and Photography. The General and the Specific, in: *Critical Inquiry*, Bd. 51, Nr. 2, 2025, 423–431, hier 425, doi.org/10.1086/732932.

20 Ankündigung von SDXL 1.0 auf der Website von Stability AI: stability.ai/news/stable-diffusion-sdxl-1-announcement (31.3.2025).

21 Daniel L. Ruderman: The Statistics of Natural Images, in: *Network: Computation in Neural Systems*, Bd. 5, Nr. 4, 1994, 517–548, hier 517.

22 Ebd., 524.

23 Vgl. Jia Deng u. a.: ImageNet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 248–255, doi.org/10.1109/CVPR.2009.5206848 (31.3.2025).

Zugespitzt ließe sich jedoch behaupten, dass die Fotografie im naturwissenschaftlichen Kontext dem Talbot'schen «Pencil of Nature»-Paradigma eines rein maschinischen Datenbildes bis heute deutlich nähersteht, als es vielen Geisteswissenschaftler*innen lieb ist.[24] Oder aber, dass sich mit dem *natural image* ein metaphysischer Bildbegriff eingenistet hat, der jegliche Bildtechnologie zwangsläufig transzendiert.

Diese Verhaftung in den Prinzipien einer mechanischen Objektivität verstärkt sich zusätzlich durch die multimodale Verknüpfung mit ‹natürlicher Sprache›, wie sie bei Text-zu-Bild-Generatoren zum Ausdruck kommt. Denn gerade das Framing fotografischer Bilder als ‹natürliche› Repräsentation legt nahe, dass sich in ihnen auch eine begriffliche ‹Selbstverständlichkeit› ausdrückt; ein Bild, das sich ‹ohne Umweg› in Sprache übersetzen lässt, weil es schlichtweg ‹offensichtlich› ist. Dass es natürlich in der Umkehrung, wie sie bei den Bildgeneratoren zur Anwendung kommt, doch nicht so einfach ist, zeigt sich vielleicht am eindrücklichsten an der gängigen Prompt-Architektur:

> A photograph of [subject] [doing something] in [setting] during [time of day], showcasing [artistic style, color palette, or visual reference], taken with [type of camera] and [type of camera lens] at [resolution, if applicable]. The image features [elements of composition, texture, or depth], evokes [emotional or cultural context], and draws inspiration from [associated artists, celebrities, or artistic movements].[25]

Die Rhetorik des Promptings zielt also nicht nur auf die präzise Festlegung bestimmter technologischer Bildgebungsmerkmale der Kamera (Kameratyp, Objektivtyp, Auflösung, Seitenverhältnis etc.), sondern ruft auch spezifische Bildkonventionen auf, die durch die Fotografie etabliert wurden (Tiefenschärfe, Bokeh, Prädikate wie *incredibly lifelike*). Die vermeintlich abwesende Kamera taucht im Prozess der Bildgenerierung dementsprechend als ein fotografisches Residuum der mechanischen Objektivität wieder auf: als die Integration einer formelhaften Rhetorik, die zwar auf eine *objektive* Darstellung ohne Eigensinn abzielt, sich dabei jedoch explizit eigensinnig fotografisch gebiert. Der ‹Blick nach außen›, durch den Sucher der Kamera, wird ersetzt durch eine konditionierte Introspektion. Das Sehen beschränkt sich somit auf einen nachträglichen Akt – eine Tätigkeit des Scannens, Filterns und vor allem Aussortierens der generierten Bilder (die nicht selten im Quartett gegen sich selbst antreten müssen) als auch zwischen jenen Bildern und einer wie auch immer gearteten fotorealistischen Präkonzeption.

Was sich also bezüglich des spezifischen Umgangs mit generierten Bildern beobachten lässt, ist, dass jene techno-logische Divergenz zwischen dem generierten und dem präkonzeptualisierten Fotorealismus nicht zwangsläufig auch eine gleichmäßige Verschiebung der zugrunde liegenden Epistemologie mit sich bringt. Mit anderen Worten: Während sich der Realismus dieser Bilder durch die Bottom-up-Logik neuronaler Netze zwar jenseits der gewohnten physikalischen Gesetze fotografischer Repräsentation und Index-Paradigma

24 Wasielewski verweist in einer anekdotischen Passage ihrer Replik auf das Problem: «At a computer vision conference I recently attended, where I was the lone humanist in the program, I addressed some of these issues to the researchers present, namely that thinking about representation matters to the kind of analysis they perform and that their assumptions of ground truth are difficult to sustain from a humanistic point of view. I was met by both bemused curiosity and shrugs». Wasielewski: Critical Response V, 424.

25 Dieses Prompt-Template geht zurück auf ein YouTube-Video auf dem Kanal des Diplom-Informatikers und «Digital Native seit 1995» Viktor Dite @TECHNiCKR: Midjourney Describe: Erstelle zu jedem Bild ein perfektes Prompt (deutsch), *YouTube*, 9.4.2023, youtube.com/watch?v=oixGeLl62So (31.3.2025), TC 00:05:30.

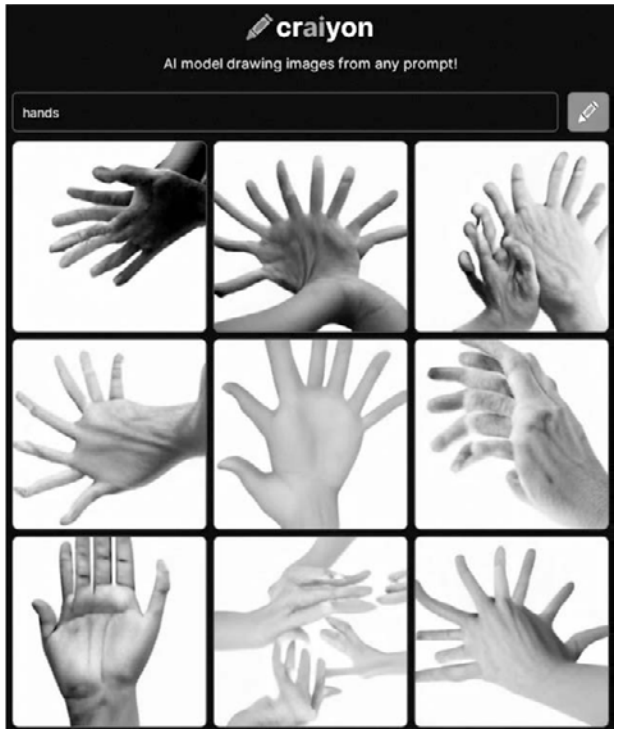

Abb. 2 **Prompt-Ergebnisse der App Crayion zu «hands»**

‹neu› formiert, besteht die ‹alte› symbolische Ordnung des fotografischen Realismus vor allem auf rhetorischer Ebene weiter. Diese Divergenz lässt sich sehr gut an einem Beispiel beschreiben, das angesichts der rasanten Entwicklung generativer Bildmodelle beinahe schon anekdotische Qualität mit sich bringt, jedoch dem Kernproblem dieser epistemologischen Verwerfung einen greifbaren Ausdruck verleiht.

III Handfeste Probleme

Kurz nach der Einführung eines Updates von DALL-E im Frühjahr 2023 arbeitete sich eine weitestgehend sarkastische Debatte an den vermeintlich offensichtlichen Schwierigkeiten ab, die generative Modelle mit der Darstellung von Händen hätten.[26] Die dort besprochenen Bilder zeigten eine Vielzahl von Anomalien, etwa unnatürlich viele oder fehlende Finger, verdrehte oder verschmolzene Gliedmaßen sowie Hände, die weitere handartige Körperstrukturen hervorbringen. Unter dem Schlagwort ‹hand problem› rief der Diskurs dabei Reaktionen hervor, die von Belustigung über Unbehagen bis hin zu Frustration reichten und sich in diversen Reddit-Threads, Memes oder dem Negativ-Prompt «no hands» niederschlugen (vgl. Abb. 2).[27]

Bemerkenswert ist, dass diesbezügliche Diskussionen oft jegliche ästhetische Überlegungen zu solchen Bildern außer Acht lassen. Anstatt die modellierte

26 Vgl. Kyle Chayka: The Uncanny Failures of A.I.-Generated Hands, in: *The New Yorker*, 10.3.2023, *newyorker.com/culture/rabbit-holes/the-uncanny-failures-of-ai-generated-hands* (31.3.2025).

27 Eine Übersicht über Ursprung und Verbreitung der KI-Hand-Memes findet sich unter AI Drawing Hands [Datenbankeintrag], *KnowYourMeme*, erstellt von Nutzer*in Philipp am 1.2.2023, letztes Update von Nutzer*in LiterallyAustin am 27.1.2025, *knowyourmeme.com/memes/ai-drawing-hands* (31.3.2025); vgl. auch den Sub-Reddit r/weirddalle, *reddit.com/r/weirddalle/comments/18475d3/why_is_ai_so_bad_at_doing_hands* (31.3.2025) sowie den *No-hands*-Meme-Generator auf der Seite Imgflip, *imgflip.com/memegenerator/113913792/No-hands* (31.3.2025).

Interpretation jener Hände als eigenständige Qualität zu befragen, wurde (und wird) die repräsentationale Abweichung als Beweis für die Unzulänglichkeit des Modells herangezogen. Diese Qualität ließe sich dabei vielleicht am ehesten mit einem räumlich-dynamischen Denkbild, wie etwa der Exzentrik einholen. Denn obwohl der fotografische Realismus die Rahmung ihrer Formgebung bestimmt – durch die Art und Weise, wie sich die physikalischen Gesetze der Optik in die Belichtungslogik oder in die spezifische Detailtreue bei der Oberflächenbeschaffenheit von Texturen einschreibt –, scheint den Bildern gleichzeitig eine fotorealistische Spezifik zu entgleiten: nämlich im repräsentativen Sinne eindeutig zu sein. Denn was im Sprechen über sie deutlich wird, ist, dass die generierten Hände einerseits ‹händisch› genug aussehen, um sie global als Hände zu identifizieren, sie jedoch gleichzeitig Bedenken hinsichtlich ihrer Repräsentationsfunktion hervorrufen.

Doch auch aus einer symbolischen Perspektive sind diese Hand-Bilder exzentrisch: weil sie genau jener totalitären Semantisierung zuwiderlaufen, die sich in die Trainingsdaten dieser Modelle einschreibt, und weil in jenen Polymorphismen ein emanzipatorisches Moment, ein *surprising change* im Zuge der Formgenese sichtbar wird. Die datenfermentierten Hände bringen etwas hervor, das sich mit Hannes Bajohr als «dumb meaning» bezeichnen lässt: eine ‹stumpfe› oder ‹dumme› Bedeutung, die sich der traditionellen Hermeneutik fotografischer Bilder entgegenstellt.[28] Eine *queerness*, die aus einem Dialog der Technik mit sich selbst heraus entsteht und die somit gerade der menschlichen Perspektive epistemologisch etwas entgegenzusetzen vermag. Denn interessant an diesen Händen ist ja gerade, dass sie sich auf visueller Ebene gegen eine fotorealistische, von aller Zweideutigkeit befreite und somit normative Repräsentation der Hand auflehnen und demgegenüber eine Art ‹residuales› Bild der Hand performen, das in seiner exemplarischen *many-to-one*-Logik eine Alternative zu einem rein repräsentativen Begriff von ‹Diversität› berührt: «We see here that we are not seen – or perhaps not seen but overseen. We are an image that machine learning sees overlaid across the total sum of the recorded».[29] Das Fotorealistische, das sich in den generativen Modellen formiert, ist demnach potenziell *non-human* in dem Sinne, dass hier tatsächlich die Emergenz einer Bildlichkeit jenseits menschlicher Perspektiven bzw. das Potenzial eines *visual world-making* möglich wird.[30]

Im tatsächlichen Umgang mit ihnen lässt sich jedoch beobachten, dass dieses analytische Potenzial meist als ‹Glitch› oder ‹Fehler› semantisch in den Abstellraum des Nonsens verbannt wird. Vor der Folie eines kategorischen Fotorealismus bleibt für die peripheren Hände nur noch das Etikett der Unplausibilität oder der Unwahrscheinlichkeit übrig. Gerade die Selbstverständlichkeit, mit der über das ‹Handproblem› als eine lediglich ‹technologische Herausforderung› gesprochen wird, beruht folglich auf der impliziten Erwartung, dass die generierten Bilder ein trennscharfes Weltbild in einer Bildwelt fotografischer Eindeutigkeit reproduzieren, die über die letzten anderthalb Jahrhunderte kollektiv eingeübt wurde.

28 Hannes Bajohr: Dumb Meaning. Machine Learning and Artificial Semantics, in: IMAGE: *The Interdisciplinary Journal of Image Sciences*, Jg. 19, Nr. 1: *Generative Imagery: Towards a ‹New Paradigm› of Machine Learning-Based Image Production*, 2023, 58–70, doi.org/10.25969/mediarep/22325.

29 Avery Slater: Critical Response IV: This Photo Does Not Exist. Generativity and the AI Gaze, in: *Critical Inquiry*, Bd. 51, Nr. 2, 2025, 416–422, hier 422, doi.org/10.1086/732925.

30 Vgl. Joana Zylinska: *Nonhuman Photography*, Cambridge (MA) 2017, 63.

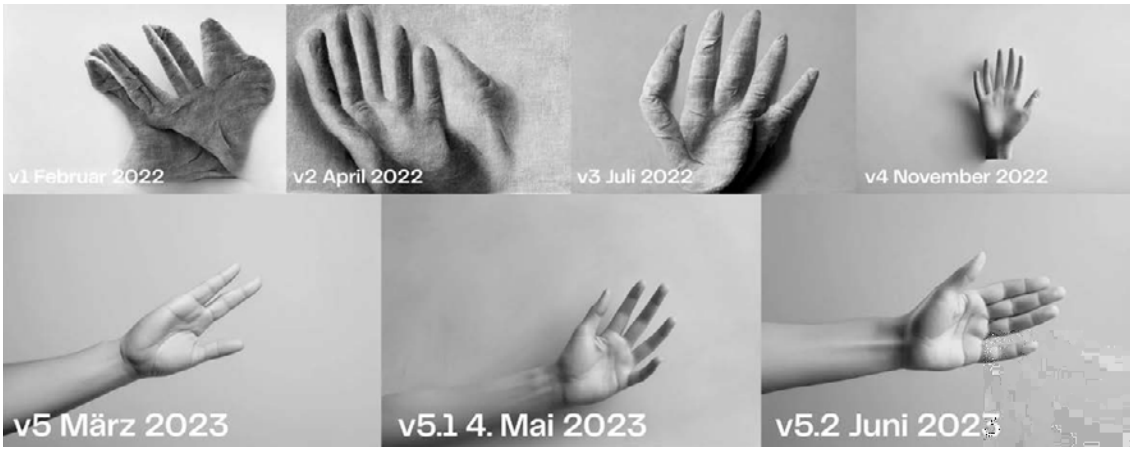

Abb. 3 **Midjourney-Entwicklung von Version 1 bis Version 5.2 bei der Darstellung von Händen**

Die unmittelbare Reaktion auf die diskursive Feedback-Schleife zum ‹Handproblem› resultierte in einem massiven *upscaling* von fotografischen Hand-Datensätzen. Darüber hinaus wurden Anpassungen vorgenommen, die das ‹Bild der Hand› durch eindeutigere Parameter stärker eingrenzen, oder um es ex negativo zu formulieren: die Mehrzahl an potenziell möglichen Händen ausschließen. Wir haben es also mit einem Normierungsprozess zu tun, der nicht nur dem *queeren* Potenzial generativer Modelle entgegenwirkt, sondern gleichzeitig einen fotorealistisch-normativen Bildbegriff der Hand zementiert, der eine Abweichung vom ‹Original› zunehmend unwahrscheinlicher macht.[31] Dass generierte Bilder einen kritischen Zeitkern haben, der von Bedeutung ist, zeigt sich dabei vielleicht recht anschaulich in einem Tableau, das im vertrauten Modus einer bildförmigen Teleologie die ‹Evolution› der KI-Hände im Durchgang der Midjourney-Updates seit 2022 darstellt (vgl. Abb. 3). Aber auch das ist letztlich eine – wenn auch weniger euphorisierende – Konsequenz einer sympoietischen Systemlogik, in der diese Bilder zwar nonhumanes Potenzial mit sich bringen, jedoch nicht außerhalb menschlicher Bezugssysteme existieren.

Denn am Ende zählt immer auch, was jenseits der offensichtlichen Feedback-Schleifen dieser chaotischen und relationalen Prozesse residual, in der ‹*rag-bag of values*› mitläuft, oder um es mit Roland Meyer zu sagen: was als ‹plausibel› erachtet wird.[32] Das zu Anfang besprochene Bildnis von Hito Steyerl ließe sich mit der skizzierten Theorie der *residual images* auch als ein durchaus realistisches Porträt in der Logik eines generativen Fotorealismus lesen, in dem jedoch weder Repräsentation noch Idealisierung, sondern vielleicht das stets verdrängte Mittelmäßige den Maßstab setzt. Welches Identifikationspotenzial würde sich in einer solchen paradigmatischen Wendung eröffnen? In jedem Fall kann es sich nicht mehr ohne Weiteres auf eine selbstverständliche ‹Anschaulichkeit› berufen. Der im Wortsinn ‹illustrative› Wert liegt vielleicht gerade im *troubling* der gewohnten Kategorien, im Aufscheinen des Obskuren und Uneindeutigen.

31 Wasielewski zeigt anhand einer Bild-Dokumentation auf, wie sie nach dem Update von DALL-E 2 im April 2022 vergeblich versucht, eine Hand mit vier Fingern bzw. eine Hand mit ‹fehlendem Finger› zu generieren. Die Ergebnisse zeigten dabei jeweils Hände mit fünf Fingern, wobei ein oder mehrere Finger im Modus des Fingerzählens gebeugt oder miteinander verschränkt waren; vgl. dies.: «Midjourney Can't Count»: Questions of Representation and Meaning for Text-to-Image Generators, in: IMAGE. *The Interdisciplinary Journal of Image Sciences*, Jg. 19, Nr. 1: *Generative Imagery: Towards a ‹New Paradigm› of Machine Learning-Based Image Production*, 2023, 71–82, hier 75, doi.org/10.25969/mediarep/22327.

32 Vgl. Roland Meyer @bildoperationen: #Platform-Realism is not an aesthetic of authenticity, but of plausibility, *Instagram*, 20.11.2023, instagram.com/p/Cz4AUOAIVpq/?hl=de&img_index=1 (31.3.2025).

Um den Trendverlauf der residualen Mittelmäßigkeit (auf welche die Generatoren rein rechnerisch abzielen) steht es jedenfalls mit Blick auf die zunehmende ästhetische Einhegung dieser Modelle nicht unbedingt gut. *Quo vadis*, generischer Fotorealismus?

Die Verantwortung dafür liegt indes weder ausschließlich bei den haldenartigen Datenkonvoluten oder der kompositorischen Logik generativer Modellarchitekturen noch bei den Geld- und Entscheidungsgeber*innen, sondern verteilt sich zusätzlich auf unzählige Mikroprozesse innerhalb einer Mensch-Maschine-Relation, in der letztlich auch alle User*innen ihr eigenes moralisches Mikro-Milieu formieren und implizit zu einer übergeordneten Autorität der Maschine beitragen. Wenn jenen Bildern ein ‹dokumentarischer Ausdruck› zu eigen ist, wie es Steyerl zu Recht nahelegt, liegt dieser vielleicht in einer neuartigen Bild-Subjekt-Konstitution begründet. Denn die Bilder sind letztlich die modellhafte Oberflächenerscheinung einer globalen Feedback-Schleife moralischer Wertesysteme, bei dem das einzelne sichtbare Bild selbst zu einem schnell entsorgbaren Durchgangsprodukt wird – ein funkenflugartiger Rest, an dem sich jedoch spontan ganze Ideologien entzünden können.