

Das Archiv als Ort der Bildproduktion: Ein Expert*inneninterview mit Roland Meyer

Erz, Hendrik

Veröffentlichungsversion / Published Version

Sonstiges / other

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

Verlag Barbara Budrich

Empfohlene Zitierung / Suggested Citation:

Erz, H. (2023). Das Archiv als Ort der Bildproduktion: Ein Expert*inneninterview mit Roland Meyer. *Soziologiemagazin : publizieren statt archivieren*, 16(1), 85-92. <https://doi.org/10.3224/soz.v16i1.06>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Das Archiv als Ort der Bildproduktion

Ein Expert*inneninterview mit Roland Meyer

geführt von Hendrik Erz

85

SozMag: *Sie beschäftigen sich mit Fotos und Kunst, allerdings nicht aus einer künstlerischen Perspektive, sondern, wie an Ihrem Buch „Operative Porträts“ [eine [Rezension](#) findet sich in Ausgabe 01/2020] sichtbar wird, aus der Perspektive der Identifizierbarkeit. Wie ist es dazu gekommen?*

Roland Meyer: Ein Interesse an Praktiken der Identifizierung hatte ich schon länger; tatsächlich stand das am Anfang meines Promotionsprojektes. In der Fotografiegeschichte ist die Polizeifotografie schon lange ein Thema. Seit den 1980er Jahren gab es, geprägt durch die Rezeption Michel Foucaults, eine ganze Reihe Arbeiten, die die Verbindung von polizeilicher, fotografischer Praxis und Porträtpraxis untersuchten. Ich denke da vor allem an Allan Sekula, John Tagg und, etwas später, Susanne Regener.

Mich interessierte im Anschluss an diese Arbeiten vor allem eine spezifisch bildwissenschaftliche Fragestellung: Was heißt es, Bilder der Identifizierung, wie Polizeifotografien oder auch Fingerabdrücke, als, mit Harun Farocki gesprochen, „operative Bilder“ zu betrachten? Als Bilder, die in standardisierten bürokratischen Operationen instrumentalisiert und lesbar gemacht werden. Der nächste Schritt war, die Geschichte dieser Operationen als Vorgeschichte der automatisierten Gesichtserkennung zu verstehen. Ich habe ich mir also die Bildgeschichte der Identifizierbarkeit seit dem 18. Jahrhundert vor dem Hintergrund gegenwärtiger Entwicklungen angeschaut: Wann fing das eigentlich an, dass man aus Bildern von Gesichtern Daten extrahiert hat, die sich dann weiterverarbeiten lassen?

SozMag: Die neueste Entwicklung dieser Praktiken scheinen nun moderne Bildgenerationsmodelle wie *Stable Diffusion*, *DALL-E* oder *Midjourney* zu sein. Was verstehen Sie unter Bildgenerierung, oder vielleicht auch Bildsynthese?

RM: Die Begrifflichkeit finde ich nach wie vor schwierig. Spricht man von Generierung, von Synthese, vielleicht auch von *Interpolation*? Und was machen eigentlich diese Tools, von denen jetzt viel die Rede ist?

86 Die aktuelle Generation von KI-Bildgenerierungstools begann mit der im April 2022 vorgestellten zweiten Version von *DALL-E*, die es faszinierend einfach möglich machte, allein mittels Texteingaben, so genannten *Prompts*, zu produzieren. Inzwischen sind weitere Programme dazugekommen, wie *Stable Diffusion* oder *Midjourney*, die inzwischen wohl die eindrucksvollsten Bilder produzieren. Rein technisch handelt es sich bei all diesen Text-to-Image-Generatoren um so genannte Diffusion-Modelle. Davor gab es andere Techniken der Bildgenerierung wie *Generative Adversarial Networks* (GAN). Man erinnert sich vielleicht an die Webseite „This Person Does Not Exist“, die fotorealistische Bilder von Gesichtern präsentiert hat, die keinen real existierenden Menschen entsprachen.

Die Frage: „Generierung oder Synthese?“, deutet das Problem daran an: Was sind eigentlich die Voraussetzungen für das,

was da passiert? Was diese Technologien benötigen, sind große Mengen von Trainingsdaten: Massen an Bildern, von denen sie „lernen“, bestimmte visuelle Muster zu reproduzieren – zum Beispiel Gesichter. Deswegen ist vielleicht Generierung schon eine problematische Deutung, weil sie suggeriert, dass da etwas aus dem Nichts geschaffen wird.

Tatsächlich *generative* Kunst gibt es bereits seit den 1960er Jahren. Diese basierte auf Programmen, die entwickelt wurden, um zum Beispiel abstrakte Kompositionen hervorzubringen. Die Bilder, die so entstanden, wurden tatsächlich rein algorithmisch generiert ohne, dass es dafür ein direktes Vorbild gebraucht hätte. Bei *Diffusion-Modellen* dagegen werden visuelle Muster bereits vorhandener Bilder statistisch ausgewertet und bilden die unverzichtbare Basis für die Produktion neuer Bilder. Das bringt viele auch zu dem Schluss, dass KI-Modelle letztlich nichts als Plagiate produzieren, sich also fremdes geistiges Eigentum aneignen.

Im Falle von *Stable Diffusion* sind es 5 Milliarden Bildtextpaare, die zu Trainingszwecken einfach aus dem Netz abgesaugt wurden. Dabei wurden Urheberrechtsfragen weitgehend ignoriert. Bilder von Künstler*innen, Illustrator*innen, Fotograf*innen und allen möglichen Kreativen dienen jetzt als Datenrohmaterial zur Produktion neuer Bilder. Nicht allein die Maschine ist es also, die mittels eines Programms etwas generiert,

ROLAND MEYER

Roland Meyer ist Bild- und Medienwissenschaftler und forscht derzeit am DFG-Sonderforschungsbereich *Virtuelle Lebenswelten* der Ruhr-Universität Bochum zu virtuellen Bildarchiven. Seine Dissertation *Operative Porträts. Eine Bildgeschichte der Identifizierbarkeit von Lavater bis Facebook* erschien 2019 bei Konstanz University Press, sein Essay *Gesichtserkennung* 2021 in der Reihe *Digitale Bildkulturen* bei Wagenbach.



sondern es handelt sich um eine statistische Synthese auf der Basis von Vorhandenem.

SozMag: *Im öffentlichen Diskurs sind derzeit zwei gegensätzliche Positionen sichtbar. Die eine Position erklärt, Diffusion-Modelle würde Kunst entwerten. Auf der anderen Seite gibt es aber Personen wie den KI-Künstler Refik Anadol, die KI verwenden, um Kunst überhaupt erst zu erschaffen. Wie stehen Sie zu dieser Debatte?*

RM: Die Frage des Werts finde ich zentral im Nachdenken darüber, was diese Modelle tun. Sie ermöglichen es nämlich von jedem Bild unendlich viele Variationen herzustellen. Man kann das eine „Inflation der Bilder“ nennen.

Diese Modelle sind aber nicht dazu in der Lage, selbst Kunst zu machen. Für Kunst braucht es – zumindest im gängigen post-

konzeptuellen Sinn der Gegenwartskunst – immer noch die Figur der Autor*in, die das, was auch immer dann als Kunst präsentiert wird, autorisiert und zur Diskussion stellt, als Teil einer individuellen oder auch kollektiven künstlerischen Praxis. Dabei kann im Prinzip – das ist seit Marcel Duchamp, Allan Kaprow und Elaine Sturtevant klar – alles zur Kunst werden: vom simplen Gebrauchsgegenstand über ein soziales Ereignis bis hin zur Aneignung bereits existierender Kunstwerke. Selbstverständlich können also auch KI-Bilder Teil einer künstlerischen Praxis werden. Was also den engeren Bereich der Gegenwartskunst angeht bedeutet KI keine Gefährdung, sondern stellt eher ein neues Medium mit neuen Möglichkeiten dar.

Die Frage ist jedoch, wie man mit diesem Medium umgeht, und das ist eine ästhetische wie eine politische Frage. Gerade Refik Anadol wäre für mich ein Beispiel

für jemanden, der diese Technologien sehr erfolgreich, aber auch äußerst affirmativ einsetzt. Anadol inszeniert die riesigen Datenmassen, die in heutige KI-Modelle einfließen als eine Art erhabenes Naturspektakel. So entsteht der Eindruck, „Künstliche Intelligenz“ in der Form, in der sie derzeit kommerziell vorangetrieben wird, sei sozusagen unsere neue Natur: etwas, dem wir nicht entkommen können, sondern wir nur andächtig bestaunen können. In den kuratorischen Texten zu Anadols Ausstellungen, wie derzeit in Düsseldorf, werden dann die ganz großen Menschheitsthemen – Einheitsfantasien von Natur und Kultur, Kunst und Wissenschaft – angesprochen, ohne dass überhaupt klar wird, was diese Technologien tatsächlich machen und auf welchen Voraussetzungen sie beruhen. Das trägt nicht zur Aufklärung bei, sondern befördert eine entpolitisierte Ästhetisierung der Technik, die von den KI-Firmen natürlich gerne gesehen wird.

Gleichzeitig gibt es viele Künstler*innen, die einen kritischen Blick auf die so genannte „künstliche Intelligenz“ werfen und dies in ihrer Arbeit thematisieren. Ich denke da zum Beispiel an Hito Steyerl, Adam Harvey oder Trevor Paglen, die auch mit KI-Kritiker*innen aus dem Bereich der Wissenschaft kooperieren, wie etwa Kate Crawford. Sie alle stellen Fragen danach, wo eigentlich diese Datensätze herkommen, welche Kategorien in das Training von Machine-Vision-Algorithmen einfließen und was für Formen

unsichtbar gemachter, häufig prekärer Arbeit die materielle Basis dafür liefern. Solche Fragen erlauben eine Kritik der KI-Bilder, die nicht dabei stehenbleibt, ästhetische Oberflächen zu bestaunen.

SozMag: *Nicht nur Künstler*innen sind von den Fähigkeiten dieser Diffusion-Modelle betroffen, sondern auch andere Bereiche wie Stock-Photography. Was sind die Konsequenzen dieser Modelle auf solche Berufsgruppen?*

RM: Das wird sich zeigen müssen. Aber klar ist: Diese Modelle ermöglichen es, sehr schnell und billig beispielsweise Illustrationen, Cover- und Plakatmotive oder auch Character Designs zu produzieren – in einer zumindest leidlich guten Qualität. Das entwertet massiv die Arbeit von Zeichner*innen, Fotograf*innen, Concept Artists für Film- und Videospiele und vielen andere Kreativen. Und weil diese Modelle ohne Rücksicht auf das Urheberrecht mit deren kreativen Produktionen gefüttert wurden, herrscht in diesen Branchen nicht nur berechtigte Sorge, sondern auch große Wut über das Vorgehen der KI-Firmen.

Allerdings sehe ich auch, dass es eine vorsichtige Neugier auf die Möglichkeiten dieser Technologien gibt. Das war zumindest mein Eindruck, als ich kürzlich auf der Pictoplasma war, einem Festival für Character Design, wo sich Illustrator*innen

und Kreative aus verwandten Bereichen austauschen. Eine kreative Nutzung dieser Technologien schien für viele dort sehr interessant. Doch insgesamt sind die Konsequenzen noch kaum absehbar.

Die plausibelste Annahme ist – wie bei vielen Fällen der Automatisierung –, dass menschliche Arbeitskraft nicht vollständig ersetzt wird, aber dass sie entqualifiziert und prekariert wird. In vielen Bereichen, wo vorher hochqualifizierte Spezialist*innen gearbeitet haben, wird es künftig eher darum gehen, Dinge auszuwählen und nachzubearbeiten, die die KI massenhaft ausspuckt. Zugleich wird der Zeit- und Kostendruck erhöht, weil man weiß, dass man mit diesen Technologien bestimmte Prozesse beschleunigen kann.

Einige versuchen, über Urheberrechtsklagen eine Form der Regulierung zu erreichen. Es gibt schon jetzt Modelle, wo zum Beispiel Stockfotograf*innen für die Nutzung ihrer Bilder als Trainingsdaten entlohnt werden. Nur sind das winzige Beträge, die in keinem Verhältnis zur Arbeitsleistung stehen. Ich erwarte daher eine ökonomische Verschiebung im Feld der Bildproduktion, die vielleicht sogar noch radikaler ausfallen wird als die, die sich in der Musikbranche vor 20 Jahren erst mit *Napster* und dann später mit Plattformen wie *Spotify* eingestellt hat.

Denn eines muss man sich klar machen: Für diese KI-Modelle hat das einzelne Bild

praktisch keinen Wert, sondern der Wert liegt im Gesamtmodell, das man aus riesigen Bildermassen extrahiert. Das ist eine ökonomische Verschiebung vom Wert des Einzelbildes, das man, wie das Bildagenturen seit Jahrzehnten machen, über Lizenzen monetarisieren kann, hin zur Gesamtheit eines Archivs, das jetzt gewissermaßen selbst zum Ort der Bildproduktion wird.

SozMag: *Was meinen Sie mit dem Archiv als „Ort der Bildproduktion“?*

RM: Solche KI-Modelle lassen sich als virtuelle Archive betrachten. Man könnte sagen, dass alle Bilder, die in das Training des Modells eingeflossen sind, gewissermaßen einen Latenzraum aufspannen, in dem man nach weiteren Bildern suchen kann. Was man da macht, würde ich, mit einer vielleicht paradox anmutenden Formulierung, als eine Form generativer Suche beschreiben. Wenn man einen Prompt eingibt, dann ist das kein Kommando, das einfach ausgeführt wird. Sowohl meine Erfahrung als auch mein Verständnis dessen, wie diese Modelle funktionieren, geht eher dahin zu sagen: Ein Prompt ist eine Art Suchbefehl.

Ähnlich, wie man in einem Bildarchiv oder einer Datenbank nach einem bereits vorhandenen Bild suchen kann, das bestimmten Schlagworten entspricht, sucht man in diesem Latenzraum möglicher Bilder nach einem Bild, das es in dieser Form noch gar nicht gibt. Diese potentiellen Bilder sind

quasi die „Lücken“ im Archiv, sie existieren „zwischen“ den bereits vorhandenen Bildern als statistische Möglichkeit.

So erklärt sich vielleicht auch das seltsame Verhältnis von Neuem und Bekannten, das diesen Bildern eigen ist. Sie sind in gewisser Weise neu und ungesehen, aber gleichzeitig rekurren sie immer auf etwas, was man schon kennt. Sie existieren letzten Endes nur in einem Möglichkeitsraum der vorgezeichnet ist von all den Bildern, die es bereits gibt. Das wird besonders anschaulich in der ästhetischen Vorliebe für eine künstlich erzeugte Nostalgie, die man derzeit in vielen KI-Bildern beobachten kann: grobkörnige Standbilder aus alten Filmen, die es nie gegeben hat; Schwarzweiß-Fotografien von Szenen, die historisch vielleicht möglich gewesen wären, aber so nie stattgefunden haben; imaginäre Vergangenheiten, die aus dem virtuellen Archiv möglicher Bilder heraus generiert wurden.

SozMag: *Das führt uns auch zu der Frage nach Biases in diesen Modellen. Wie mit allen KI-Technologien gibt es auch bei Bildgeneratoren den Vorwurf, diese hätten einen starken pro-westlichen Bias. Welche Probleme sehen Sie in dieser Hinsicht, und gibt es bereits Möglichkeiten, diesen entgegenzuwirken?*

RM: Diese KI-Modelle sind im Prinzip Klischeeverstärker, die die visuellen Muster und Stereotypen unserer digital vernetzten visuellen Kultur, internalisieren, verdichten

und reproduzieren. Dadurch sind sie extrem anfällig für Bias auf unterschiedlichsten Ebenen. Da ein Großteil der Trainingsbilder und -daten, soweit man das nachvollziehen kann, aus dem globalen Norden und vor allem aus den USA stammt, weisen die KI-Bilder bestimmte Tendenzen auf. Im Standardmodus sind es etwa überwiegend *weiße*, normschöne Gesichter, die diese Software produziert. Auch Geschlechterrollen sind häufig stereotyp. Das betrifft nicht zuletzt Zuordnungen von Geschlecht und Beruf: Beim Prompt „CEO“ wird dann in der Regel ein *weißes*, männliches Gesicht und bei „assistant“ ein weibliches Gesicht synthetisiert, auch dieses häufig *weiß*, manchmal auch asiatisch.

Es gibt Versuche der Firmen, diesem Bias entgegenzuwirken. Doch findet das zum Teil gar nicht auf der Ebene der Modelle statt, sondern auf der Ebene der Interfaces. So schummelt DALL-E seit letztem Sommer bei manchen Prompts bestimmte Wörter hinzu, ohne dass die User*innen es merken, zum Beispiel „female“ oder „black“, um so eine Diversität im Ergebnis zu produzieren, die sonst nicht da wäre. Das löst das Problem jedoch nicht, sondern wirkt bloß kosmetisch.

SozMag: *Ein weiteres aktuelles Thema ist die Frage nach Bildgeneratoren und der Wahrheit. Es gibt Diskussionen darüber, welchen Bildern man online noch vertrauen kann; beispielsweise der „Pontiflex“ oder Angela Merkel und Barack Obama am Strand. Welchen Einfluss*

*haben diese Bildgeneratoren vor allen Dingen mit Blick auf Akteur*innen, welche online Unruhe stiften wollen, Leute verunsichern oder die Arbeit von Journalist*innen schwerer machen?*

RM: Auf Twitter gab es auf den KI-Papst Reaktionen wie: „Das Zeitalter der Wahrheit ist vorbei.“ Wenn man das liest, denkt man sich: „Wann soll denn das gewesen sein?“ Vor 30 Jahren konnte man ungefähr die gleichen Sätze angesichts von Photoshop lesen. Und fotografische Retusche gab es schon lange davor. Es gab immer schon Bildmanipulationen.

Für alle, die professionell mit Bildern umgehen, sei es in Bildredaktionen oder im Kontext forensischer Ermittlungen, hing die Wahrheit der Bilder vielleicht noch nie am einzelnen Bild. Um die Authentizität und Beweiskraft von Bildern zu sichern, ist es vielmehr notwendig, dass man Bilder mit anderen Bildern und Daten abgleicht: Ist es überhaupt plausibel, dass dieses Bild zum angegebenen Zeitpunkt dort aufgenommen wurde, wo es angeblich aufgenommen wurde? Stimmen die Lichtverhältnisse, die Wetterverhältnisse, gibt es andere Bilder des gleichen Ereignisses, können wir weitere Quellen finden, die das Gezeigte bestätigen? Es gibt ein ganzes Set von eingeführten Verfahren der Absicherung von Evidenz – bis hin zu bildforensischen Techniken, die tief in die Datenstrukturen der Bilder vordringen und dort nach Spuren der Manipulation

suchen. Und diese Verfahren werden sich weiter ausdifferenzieren und aktualisieren angesichts der KI-Bilder.

Was man jedoch beobachten kann – und ich auch an mir selbst beobachte –, ist eine Veränderung der spontanen Reaktionen auf ein Bild, mit dem man online konfrontiert ist. Der Standardmodus scheint mir mittlerweile die Skepsis, ja der Verdacht gegenüber den Bildern zu sein. Ein interessantes Beispiel dafür war letztes Jahr eine Aufnahme der Regenbogenflagge vor dem Kanzleramt, bei der ein FAZ-Redakteur auf Twitter insinuiert hat, sie wäre manipuliert, weil nur diese Regenbogenflagge im Wind flatterte, während die anderen Flaggen gerade schlaff herunterhingen. Ohne zu bedenken, dass, wer auch immer das fotografiert hat, natürlich genau den Moment abgewartet hat, wo diese Flagge tatsächlich gerade im Wind flattert. Wenn also ein journalistischer Profi in diesen Modus des Verdachts gegenüber dem Bild verfällt, und das damals auch noch weitgehend unbeeindruckt von den Möglichkeiten der KI-Bildgenerierung, zeigt das nicht zuletzt, wie anschlussfähig eine solche Bilderskepsis in den sozialen Medien ist – und dass sie, wie in diesem Fall, immer gut dafür ist, ein paar Klicks zu generieren.

SozMag: *Zum Abschluss noch eine Frage zur Zukunft der Bildgeneration: Welche Weiterentwicklung dieser Modelle verfolgen Sie besonders gespannt?*

RM: Das Interessanteste scheint mir derzeit im Feld bewegter Bilder zu passieren, etwa Text-to-Video, also die Synthese kurzer Szenen mittels Prompts. Noch bemerkenswerter erscheinen mir die Möglichkeiten der Nachbearbeitung von Videos, die sich eröffnen, wenn Videosequenzen auf ihre basale visuelle Struktur reduziert werden und mittels Image-Prompt mit ganz neuen, KI-generierten Inhalten, Looks und Atmosphären gefüllt werden. So kann man zum Beispiel schon jetzt eine alltägliche Straßenszene nehmen und diese allein auf Grundlage eines entsprechenden Bildes in eine Cyberpunk-Science-Fiction-Landschaft verwandeln. Ich denke, dass solche Techniken zunehmend in alle möglichen Formen von Filtern einfließen werden, auf Social-Media-Plattformen wie Instagram und TikTok, aber auch in den Foto-Apps unserer Smartphones. Und das kann sehr auffällige und dramatische Transformationen bedeuten, aber auch subtile Formen der Optimierung, Verfremdung und Verschönerung. All das zielt auf eine ästhetische Neukonfigurierung unserer visuellen Umwelt, die ich faszinierend, aber auch ein wenig gruselig finde.

SozMag: Vielen Dank für das Gespräch.

LITERATUR

- Crawford, K. (2021). *Atlas of AI*. Yale University Press.
- Farocki, H. (2004). Phantom Images. *Public* (29). <https://public.journals.yorku.ca/index.php/public/article/view/30354>
- Paglen, T. (2018, 8. Dezember). *Invisible Images (Your Pictures Are Looking at You)*. <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/>
- Regener, S. (1999). *Fotografische Erfassung: Zur Geschichte medialer Konstruktionen des Kriminellen*. Fink.
- Sekula, A. (1986). Der Körper und das Archiv. In H. Wolf (Hg.), *Diskurse der Fotografie: Fotokritik am Ende des fotografischen Zeitalters*, Bd. 2. Suhrkamp.
- Steyerl, H. (2018). *Duty Free Art: Kunst in Zeiten des globalen Bürgerkriegs*. Diaphanes.
- Tagg, J. (1988). *The Burden of Representation. Essays on Photographies and Histories*. Palgrave Macmillan.

Das Interview wurde im Juni 2023 von **Hendrik Erz** online geführt und von **Annabell Lamberth** lektoriert.