Zeitschrift für Medienwissenschaft, Jg. 18, Heft 34 (1/2026), https://doi.org/10.14361/zfmw-2026-180104. Published by transcript Verlag.

LISA REIN

SCHÖNE NEUE WELT

Automatisierte Bildbewertung zwischen «subfaces» und «surfaces» von Plattformen

Die auf Konditionierung im Flüsterton basierende Gesellschaft in Aldous Huxleys dystopischem Roman *Schöne neue Welt* (*Brave New World*) von 1932 gleiche, so der Anglist Jerome Meckier, einem «multinationalen Konzern, dessen Religion im Zwangskonsum besteht», und werde von einem «Weltcontroller» beherrscht, der «halb Diktator, halb CEO ist».[1] In Anbetracht der gegenwärtigen spätkapitalistischen Verhältnisse, in denen es möglich ist, dass Politiker zu Plattformeigentümern (Trump), Plattformeigentümer zu Politikern (Musk) und beide zu politischen Akteuren innerhalb eines «neuen Faschismus»[2] werden können, erscheint der Roman überraschend aktuell. Auch wenn Huxley mit seinem Gesellschaftsentwurf keine der digitalen, plattformkapitalistischen Infrastrukturen unserer Zeit antizipiert, zeichnet er doch eine Welt, in der die Bedürfnisbefriedigung als höchstes Ziel durch chemische Substanzen, Sex und insbesondere die Dauerbeschallung mit «tanzenden Bilder[n] aus der Telebox» und «Fühlorama-Schauen» herbeigeführt wird.[3] Die Bewohner*innen der ‹schönen neuen Welt› leben in einer Aufmerksamkeitsökonomie, in der die Frage, welche Bilder zirkulieren, von höchstem Stellenwert ist. Dies wird in einer Szene deutlich, in der die Affekte von Kleinkindern im «Konditionierungstrakt» bei der Betrachtung von Bilderbüchern mit Elektroschocks gezielt geformt und automatisiert werden.[4]

In dem von vernetzten und verteilten Bildern geprägten Plattformkapitalismus unserer Gegenwart haben Bilder ähnlich hohen Stellenwert und auch hier unterliegt ihre Betrachtung und Bewertung Verfahren der Automatisierung – wenn auch etwas anders gelagert: Auf Basis von Machine-Learning-Systemen werden Bilder zunehmend hinsichtlich ihres ‹ästhetischen Werts› klassifiziert und dementsprechend in ihren Sichtbarkeiten reguliert. Die Beurteilung, welche Bilder als schön, ästhetisch und damit wertvoll und verwertbar gelten, wird in Plattformumgebungen vermehrt an algorithmische Systeme ausgelagert. Die automatisierte Bewertung von Bildern anhand von sogenannten *aesthetics ratings* durchzieht plattformkapitalistische

1 Jerome Meckier: Onomastic Satire: Names and Naming in Brave New World, in: ders.: *Aldous Huxley: Modern Satirical Novelist of Ideas. A Collection of Essays by Jerome Meckier*, hg. v. Peter Edgerly Firchow u. Bernfried Nugel, New Brunswick, London 2006, 185–224, hier 193, Übers. LR.

2 Rainer Mühlhoff: *Künstliche Intelligenz und der neue Faschismus*, Ditzingen 2025.

3 Aldous Huxley: *Schöne Neue Welt. Ein Roman der Zukunft*, übers. v. Uda Strätling, Frankfurt/M. 2018, 177, 189.

4 Vgl. ebd., 26–30.

Medienumgebungen des alltäglichen Gebrauchs, wie die «curatorial assemblages» von Social-Media-Feeds,[5] und ist damit zunehmend an der Strukturierung von Bildordnungen und Blickregimen der digitalen Gegenwart beteiligt. Wie kommen diese zunehmend algorithmisch gesteuerten Bewertungen zustande und welche Konsequenzen haben sie? Die Voraussetzungen für die algorithmenbasierte Bildbewertung bilden menschlich-maschinelle Gefüge der Klassifizierung, innerhalb derer Bildbeurteilungen akkumuliert und quantifiziert werden. Mein Beitrag untersucht diese Interface-Prozesse anhand der Datenbank LAION-Aesthetics, die exemplarisch für das sogenannte *automatic aesthetic assessment* steht, einen Teilbereich der Computer Vision. Es zeigt sich, dass entgegen beliebter Behauptungen aus der Machine-Learning-Industrie bei Weitem keine universelle oder objektive ästhetische Bildbewertung stattfindet, sondern algorithmische Klassifizierungen eigenmächtig Visualitäten hervorbringen, die maßgeblich davon abhängen, wie sich die zugrunde liegenden Datenbanken und Modelle konstituieren. Die algorithmische Bewertung von Bildern ist als kontextabhängige, menschlich-maschinelle Assemblage der Bildklassifizierung zu verstehen. Im Sinne der kritischen Interface Studies betrachte ich LAION-Aesthetics als performative Aushandlungszone, in der «menschliche und mehr-als-menschliche Operationen fluide ineinandergreifen und zunehmend integrale Bestandteile der Lebenswirklichkeit werden».[6]

Das von Frieder Nake eingeführte Begriffspaar *surface/subface*[7] verwende ich angelehnt an Jan Distelmeyers Vorschlag zur Unterscheidung der im Alltag sichtbaren Anwendungsoberflächen (*surfaces*) einerseits und der für Lai*innen schwer zugänglichen, aber nichtsdestotrotz für plattformkapitalistische Operationen unverzichtbaren tieferen Schichten (*subfaces*) andererseits.[8] Auf dieser Grundlage verorte ich *aesthetics ratings* im Kontext von plattformkapitalistischen Ranking- und Bewertungsumgebungen und diskutiere, inwiefern es sich hier weniger um ästhetische Analysen als um eine numerische Vermessung und Präskription von Geschmack handelt. Es wird deutlich, dass die auf Social-Media-Plattformen habitualisierte Bewertung von Bildern tief in die *subfaces* von gegenwärtigen visuellen digitalen Kulturen, etwa von Bilddatenbanken und großen Bildmodellen, hineinreicht. Und dass andersherum die – für Nutzer*innen größtenteils unsichtbaren – Operationen algorithmisch automatisierter Bildklassifizierung bis weit in die Plattform-*surfaces*, ihre allgegenwärtigen Bildordnungen und «regimes of recognition»[9] vordringen.

Algorithmische Bildklassifizierung in Social-Media-Feeds

Wie eine 2024 veröffentlichte Studie von Informatiker*innen der University of Wisconsin-Madison gezeigt hat, setzt Instagram bereits im Upload-Vorgang eine automatisierte Bildklassifizierung ein, welche die hochzuladenden

5 Vgl. Christoph Bareither, Sabine Wirth: Social Media Feeds as Curatorial Assemblages. A Conceptual Framework, in: *Media, Culture & Society*, Bd. 48, Nr. 1, 2026, 73–88, hier 74, doi.org/10.1177/01634437251360372.

6 Jan Distelmeyer, Timo Kaerlein, Sabine Wirth: Interfaces | Plattformen. Einleitung in den Schwerpunkt, in: *Zeitschrift für Medienwissenschaft*, Jg. 18, Nr. 34 (1/2026): *Interfaces | Plattformen*, 10–20, hier 14 (in diesem Heft).

7 Vgl. Frieder Nake: The Disappearing Masterpiece: Digital Image & Algorithmic Revolution, in: Mario Verdicchio u. a. (Hg.): *CoAx 2016. Proceedings of the Fourth Conference on Computation, Communication and X*, Bergamo 2016, 12–27, hier 13.

8 Vgl. Jan Distelmeyer: *Kritik der Digitalität*, Wiesbaden 2021, 90.

9 Benjamin N. Jacobsen: Regimes of Recognition on Algorithmic Media, in: *New Media & Society*, Bd. 25, Nr. 12, 2023, 3641–3656, doi.org/10.1177/14614448211053555.

10 Vgl. Jack West u. a.: A Picture is Worth 500 Labels: A Case Study of Demographic Disparities in Local Machine Learning Models for Instagram and TikTok, in: *Proceedings of the 2024 IEEE Symposium on Security and Privacy*, San Francisco 2024, 1–18, hier 13, doi.org/10.1109/SP54263.2024.00203.

11 Vgl. Bareither, Wirth: Social Media Feeds as Curatorial Assemblages, 80.

12 Vgl. z. B. Tarleton Gillespie: *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, New Haven 2018; Sarah T. Roberts: *Behind the Screen. Content Moderation in the Shadows of Social Media*, New Haven 2019.

13 Jacobsen: Regimes of Recognition on Algorithmic Media, 3642.

14 Vgl. Instagram: Helping Creators Find New Audiences [Beitrag auf Unternehmensblog], *Instagram for Creators*, 30.4.2024, creators.instagram.com/recommendations-and-originality?locale=en_US (29.9.2025). Dieser Blog-Post zu Instagrams Empfehlungsalgorithmen verdeutlicht, dass für das neue, mehrstufige Ranking-Verfahren Inhalte algorithmisch sortiert werden und z. B. als «low-quality» oder «political» eingestufte Inhalte Nutzer*innen nicht empfohlen werden.

Bilder und Videos unter anderem in Hinblick auf ihren ‹ästhetischen Wert› analysiert (vgl. Abb. 1).[10] Dabei wird jedes potenziell hochzuladende Bild oder Video mit einem *aesthetics rating* auf einer Skala von null bis eins versehen, und es ist davon auszugehen, dass Social-Media-Plattformen wie Instagram und TikTok solche Ratings als messbaren Indikator für das Kuratieren der Feeds verwenden. Sie sind Teil des ‹inhaltsbasierten Filterns›, das neben dem bekannten ‹kollaborativen Filtern› eine wachsende Rolle für die plattformseitige Gestaltung der Nutzer*innen-Feeds von Social-Media-Plattformen spielt.[11] Darunter fällt längst nicht mehr nur die begründete Löschung einzelner Bilder und Videos, die vor allem in den 2010er Jahren unter dem Stichwort *content moderation* breit diskutiert wurde,[12] sondern eine umfassende automatisierte Auswertung *aller* Bilder und Videos im Feed hinsichtlich möglicher Bildsujets und -themen, aber auch in Hinblick auf bildästhetische Kategorien. Der Social-Media-Forscher Benjamin Jacobsen konstatiert: «Rather than merely facilitating more image-sharing among people, social media platforms are increasingly seeking to understand the content of images on a more granular level.»[13]

Abb. 1 **Instagram Vision Model im Einsatz, inkl. Tag *aesthetics_rating* (aus West u. a. 2024)**

Aus Sicht der Plattformbetreiber*innen dient diese ausdifferenzierte Form der Bildanalyse einerseits dazu, Nutzer*innen-Interessen noch genauer zu identifizieren und die personalisierte Ansprache weiter zu perfektionieren, wie sich ein im Frühling 2024 veröffentlichtes Update von Instagram zu den plattformeigenen «recommendation rankings» unschwer erkennen lässt.[14] Andererseits regulieren Plattformen per (teil-)automatisierter Bildbewertung auch auf globaler Ebene Sichtbarkeiten, die über die einzelnen Nutzer*innen-Feeds hinaus wirksam werden. So geht aus im November 2024 geleakten internen Dokumenten von TikTok hervor, dass die Plattformbetreiber*innen als Reaktion auf ein – aus ihrer Sicht – «high volume of […] not attractive subjects» im «For You»-Feed Anpassungen an Filter-Algorithmen vorgenommen hatten, welche die Sichtbarkeiten jener Subjekte verringern sollten.[15]

Es ist kein Geheimnis mehr, dass unter den hier adressierten Algorithmen relationale Gefüge zu verstehen sind, in denen verschiedene menschliche

15 Vgl. Bobby Allyn, Sylvia Goodman, Dara Kerr: TikTok Executives Know about App's Effect on Teens, Lawsuit Documents Allege, in: NPR, 11.10.2024, npr.org/2024/10/11/g-s1-27676/tiktok-redacted-documents-in-teen-safety-lawsuit-revealed (26.11.2024).

und maschinelle Prozesse zusammenlaufen.[16] Diese bestehen, wie im Fall von TikToks mehrstufiger Moderations-Infrastruktur, nach wie vor zu großen Teilen aus manueller Datenarbeit, verrichtet von Millionen von prekär beschäftigten ‹Moderator*innen›.[17] Gleichzeitig wird die Auswertung von Bildern, wie das Beispiel von Instagrams Upload-Algorithmus zeigt, immer mehr automatisiert – insbesondere seitdem sich sogenannte Large Image Models, die nach dem Motto «crawl over curate» auf großen und manuell wenig kuratierten Datenmengen sowie selbstlernenden Algorithmen basieren,[18] zunehmend durchsetzen. Jedoch basiert auch diese automatisierte Klassifizierung letztlich auf vorgelagerten menschlichen Bildbewertungen, die über den Umweg von Modellen und Datenbanken wirksam werden. Mit dem vorliegenden Beitrag zeige ich anhand eines konkreten Beispiels, wie sich menschliche und maschinelle Klassifizierungen in den *subfaces* von Plattformumgebungen überlagern und anschließend auf Social-Media-Feeds performativ werden. Dabei greife ich mit LAION-Aesthetics auf eine Open-Source-Datenbank zu, die hier als Platzhalter für die proprietären – und damit uneinsehbaren – algorithmischen Infrastrukturen von Plattformen wie Instagram herhalten muss. Dies erweist sich insofern als gerechtfertigt, als die Befunde der eingangs erwähnten Studie zu Instagrams *aesthetics rating* oder das geleakte Google-Memo[19] vom Mai 2023 stark darauf hindeuten, dass die proprietären Technologien der großen Plattformen sehr ähnlich funktionieren.

«Screenwalking» in den Archipelen der Plattformen

Ich komme in diesem Beitrag dem Aufruf der kritischen Interface Studies nach, Prozesse computerbasierter Normierung, die – über sichtbare (User) Interfaces hinaus – konstitutiv für die plattformkapitalistische Gegenwart sind, kritisch zu diskutieren.[20] Dabei verstehe ich Interfaces als «konzeptionelle Kippfigur[en] digitaler Medien»,[21] in denen zwei oder mehr Komponenten digitaler Infrastruktur, wie Hardware, Software, Protokolle, Code, Algorithmen, Nutzer*innen und weitere Formationen, sich begegnen, vermischen und gegenseitig transformieren.[22] Inspiriert von Femke Sneltings Anwendung von Édouard Glissants «poetics of relation» verstehe ich Interfaces dabei als *archipelagos* (Archipele), als Übergangsräume, in denen verschiedene Materialitäten und Zustände immer wieder aufeinandertreffen und durch fließende, mitunter auch gewaltsame Verschiebungen ineinander übergehen.[23] Das Bild des Archipels verdeutlicht, dass die repetitive, ständig von Neuem performativ werdende Übersetzungsbewegung von Interfaces nicht als einmal ablaufender und dann abgeschlossener Prozess zu begreifen ist. Vielmehr macht die Auseinandersetzung mit plattformisierten, automatisierten Infrastrukturen der Bildbewertung deutlich, dass es sich um rekurrierende Feedbackschleifen handelt, in denen sich *sub-* und *surfaces* gegenseitig bedingen und immer wieder neu hervorbringen.

16 Vgl. Tobias Matzner: *Algorithms: Technology, Culture, Politics*, London 2023, 6.

17 Vgl. Chris Köver, Markus Reuter: TikTok: Gute Laune und Zensur, *netzpolitik.org*, 23.11.2019, netzpolitik.org/2019/gute-laune-und-zensur (25.9.2025).

18 Abeba Birhane, Vinay Uday Prabhu, Emmanuel Kahembwe: Multimodal Datasets. Misogyny, Pornography, and Malignant Stereotypes, *arXiv*, 5.10.2021, 1–33, hier 5, doi.org/10.48550/arXiv.2110.01963.

19 In dem geleakten Memo führt ein*e Google-Mitarbeiter*in eindrücklich aus, dass Open-Source-KI-Technologien mit denen von Google und OpenAI gleichauf sind. Vgl. Dylan Patel, Afzal Ahmad: Google «We Have No Moat, And Neither Does OpenAI», *Semianalysis*, 4.5.2023, newsletter.semianalysis.com/p/google-we-have-no-moat-and-neither (5.12.2025).

20 Vgl. Distelmeyer: *Kritik der Digitalität*, 95.

21 Sabine Wirth: *Formierungen des Interface. Zur Mediengeschichte und -theorie des Personal Computing*, Bielefeld 2024, 77.

22 Vgl. ebd., 74.

23 Vgl. Femke Snelting: Other Geometries, in: *transmediale journal*, Nr. 3: *Affective Infrastructures*, hg. v. Daphne Dragona, 31.10.2019, archive.transmediale.de/content/other-geometries (18.1.2026).

In Erweiterung der in den App Studies und Interface Studies bekannten Walkthrough-Methode, die häufig für kultur- und medienwissenschaftliche Analysen von digitalen Bildschirmartefakten verwendet wird,[24] setze ich ein Verfahren ein, das ich als *screenwalking* bezeichnen möchte und das es mir erlaubt, auf Bildschirmspaziergängen in die Tiefen der Datenbanken einzutauchen. Diese kursorischen Streifzüge ermöglichen ein assoziatives, teils autoethnografisches Vorgehen, das zufällige oder von spontanen Affekten geleitete Entdeckungen hervorbringt, die das Close Reading von Bilddatenbanken leiten und rahmen können. Im Zusammenspiel zwischen den so entstehenden qualitativen Beobachtungen, einer gründlichen Lektüre der informatischen Veröffentlichungen zu den jeweiligen Datenbanken und den bereits geleisteten Arbeiten der Critical Dataset Studies[25] wird eine kritische Analyse der algorithmischen Bildklassifizierung möglich. Damit erprobe ich die Umsetzung medienarchäologischer Ansätze, wie sie Antonio Somaini zur Erforschung maschineller Bildtechnologien vorgeschlagen hat.[26] Birgit Schneider skizziert in ihrem Aufsatz «Computersehen» hierzu zwei konkrete Möglichkeiten, wenn sie dafür plädiert, präzise Analysen algorithmischer Infrastrukturen in Form einer «kritische[n] Historiographie und Epistemologie» sowie in der Erprobung experimenteller, explorativer Methoden durchzuführen.[27] Das Anliegen meines Beitrags ist es, diese beiden Vorhaben zu kombinieren und damit der «Mythisierung einer gottgleichen ‹Künstlichen Intelligenz›»[28] eine pointierte Untersuchung spezifischer algorithmischer *subfaces* entgegenzustellen.

«Automatic aesthetic assessment» und die Bilddatenbank LAION

Das Feld des *automatic aesthetic assessment*, die automatisierte Bewertung von Bildern anhand einer numerischen Skala, ist ein Teilbereich der sogenannten Computer Vision, also des Gebiets des Maschinellen Lernens, das sich seit Mitte des 20. Jahrhunderts der automatisierten Klassifizierung von Bildern widmet. Das informatische Forschungsgebiet der Computer Vision wurde ab den späten 2000er Jahren stark weiterentwickelt, als eine Reihe von Entwicklungen zusammenfielen: die Steigerung von Prozessorkapazitäten, die Wiederbelebung von ‹Deep Learning›-Verfahren, die Verbreitung von Smartphones, die massenhafte Verfügbarkeit digitaler Bilder auf Social-Media-Plattformen wie Flickr und Facebook sowie neue Formen verteilter Datenarbeit.[29] Auch die automatisierte Bewertung und Optimierung von Bildern wird seit ca. 2010 stark vorangetrieben und ist in Form verschiedener Anwendungen mittlerweile fester Bestandteil gegenwärtiger digitaler Bildkulturen, wo sie insbesondere alltägliche Begegnungen mit Bildern formt. Außer beim Kuratieren von Social-Media-Feeds findet sie Verwendung bei der generativen Erzeugung synthetischer Bilder, bei Filter-Apps und Smartphone-Kamera-Anwendungen,[30] aber auch bei Rankings für Stockfoto-Plattformen[31] oder bei Bilder-Suchmaschinen wie der Google-Bildersuche.[32]

24 Vgl. Ben Light, Jean Burgess, Stefanie Duguay: The Walkthrough Method. An Approach to the Study of Apps, in: *New Media & Society*, Bd. 20, Nr. 3, 2018, 881–900, doi.org/10.1177/1461444816675438.

25 Für einen Überblick zu den Critical Dataset Studies vgl. die Webseite des Forschungsprojekts «Knowing Machines», *knowingmachines.org* (28.9.2025)

26 Antonio Somaini: Film, Media, and Visual Culture Studies, and the Challenge of Machine Learning, in: *NECSUS. European Journal of Media Studies*, Jg. 10, Nr. 2, 2021, 49–57, hier 54, doi.org/10.25969/mediarep/17289.

27 Vgl. Birgit Schneider: Computersehen. Elemente einer Medienarchäologie der Computer Vision und ihrer Sehstörungen, in: Martin Huber, Sybille Krämer, Claus Pias (Hg.): *Wovon sprechen wir, wenn wir von Digitalisierung sprechen? Gehalte und Revisionen zentraler Begriffe des Digitalen*, Bayreuth 2020, 197–226, hier 221. Dies ist auch der einzige mir bekannte Text innerhalb der deutschsprachigen Medienwissenschaft, der sich mit *aesthetics ratings* befasst.

28 Ebd.

29 Vgl. Christoph Engemann: Rekursionen über Körper. Machine Learning-Trainingsdatensätze als Arbeit am Index, in: ders., Andreas Sudmann (Hg.): *Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz*, Bielefeld 2018, 247–268, hier 249, 254.

30 Vgl. Winfried Gerling: Das Bild als Wahrscheinlichkeit, in: Olga Moskatova, Laura Katharina Mücke (Hg.): *Bild | Kanäle. Zur Theorie und Ästhetik vernetzter Medienkultur*, Würzburg 2024, 39–70, hier 49.

31 Vgl. Schneider: Computersehen, 197.

32 Vgl. Leonardo Impett: Computation and Beauty [Beitrag auf Online-Plattform], *Unthinking Photography*, 28.10.2024, unthinking.photography/articles/computation-and-beauty (3.2.2025).

Wie aus internen Anleitungsdokumenten hervorgeht, werden Datenarbeiter*innen, die Trainingsmaterial für Suchmaschinen-Algorithmen aufbereiten, beispielsweise aufgefordert, Ergebnisse von Bildersuchen anhand einer Skala von «highly satisfying» bis «fails to satisfy» einzuteilen.[33] Dabei wird «highly satisfying» als «extremely helpful, beautiful, inspirational, or visually appealing» definiert, wohingegen das Label «fails to satisfy» aus verschiedenen Gründen zugewiesen wird: So kann «extremely low image quality» genauso zu einer schlechten Bewertung führen wie «unpleasant or upsetting content», oder die einfache Tatsache, dass es sich um ein pornografisches Bild handelt – «unless the query clearly indicates the user is seeking that type of content». Hier deutet sich bereits an, mit welchen komplexen bildpolitischen Fragen die quantifizierende Bewertung von Bildern einhergeht, auf die ich weiter unten konkreter eingehen werde. Zunächst stellt sich allerdings die Frage, wie Ästhetik durch das Feld des *automatic aesthetic assessment* konstruiert wird und wie die Klassifizierung der ‹schönen neuen Welt› bzw. der schönen Welt ‹neuer› Medien konkret funktioniert. Das zeige ich im Folgenden anhand von LAION-Aesthetics, einer Datenbank aus der LAION-Familie.

LAION, kurz für Large-scale Artificial Intelligence Open Network, ist der Name für eine Serie von Bilddatenbanken, die von dem in Deutschland eingetragenen Verein LAION e.V. als semikommerziellem Open-Source-Projekt in Zusammenarbeit mit (v. a. deutschen) Universitäten und finanzieller Unterstützung von Huggingface und Stability AI entwickelt worden ist.[34] Alle LAION-Datenbanken bestehen vor allem aus zwei Dateneinheiten: Bildern und Texten. Mit den Datenbanken werden sogenannte Large Image Models trainiert, die dann über ein gewisses semantisches Wissen verfügen, also Beziehungen zwischen Bildern und Texten erlernen, die im Anschluss für Bildgenerierungen – unter anderem von dem bekannten Text-to-Image-Generator Stable Diffusion – genutzt werden. In LAION-5B, der größten und prominentesten Version von LAION, die 2022 erstmals veröffentlicht wurde, liegen fünfeinhalb Milliarden Bild-Text-Paare, die aus dem Internet gecrawled wurden. Es handelt sich also um Bilder, die im World Wide Web zwischen 2008 und 2022 veröffentlicht wurden, und die dazugehörigen Alt-Texte.[35]

Wie kritische Informatiker*innen, Künstler*innen und Geisteswissenschaftler*innen seit dem Release der ersten LAION-Version (LAION-400M, 2021) betont haben, ist der Datensatz von einer Vielzahl problematischer Verkürzungen und Verzerrungen gekennzeichnet. So befinden sich erstens «troublesome and explicit images and text pairs of rape, pornography, malign stereotypes, racist and ethnic slurs, and other extremely problematic content» sowie Bilder, die ohne Zustimmung der Abgebildeten veröffentlicht wurden, in LAION.[36] Zweitens sind die verwendeten Alt-Texte, anders als von den LAION-Entwickler*innen behauptet, selten zuverlässige Beschreibungen der

33 Diese Dokumente liegen mir vor, können aber aufgrund von Vertraulichkeitsvereinbarungen hier nicht näher benannt werden. Es handelt sich um Anleitungen in PDF-Form, die in Deutschland beschäftigte Datenarbeiter*innen erhalten, um Trainingsdaten für die Suchmaschinen-Algorithmen eines namentlich nicht genannten Unternehmens zu erstellen.

34 Vgl. Christoph Schuhmann u. a.: LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models, *arXiv*, 16.10.2022, doi.org/10.48550/arXiv.2210.08402.

35 Vgl. LAION e. V.: Releasing Re-LAION 5B. Transparent Iteration on LAION-5B with Additional Safety Fixes [Beitrag auf Unternehmensblog], LAION, 30.8.2024, laion.ai/blog/relaion-5b (11.3.2025).

36 Vgl. Birhane, Prabhu, Kahembwe: Multimodal Datasets, 1.

Bilder. Die meisten der in den 1990er Jahren als Alternativ-Beschreibungen für blinde Menschen konzipierten Alt-Texte sind heute darauf ausgerichtet, Bilder und Webseiten für algorithmische Sortiersysteme wie Google Page Rank lesbar und adressierbar zu machen. Sie dienen in der Regel der Verwertbarmachung der Bilder im Datenkapitalismus und enthalten oftmals klischeehafte Beschreibungen, unzutreffende Bildinterpretationen oder kryptische, für algorithmische Rankingsysteme optimierte Maschinensprache.[37] Drittens befinden sich in LAION überproportional viele Bilder von stark frequentierten Webseiten wie Shopify oder Pinterest, die Alt-Texte gezielt zur Umsatzsteigerung einsetzen.[38]

Trotz der eindeutigen Biases brüstet sich Stability AI auf dem Webseiten-Eintrag zu Stable Diffusion damit, das Modell würde die gesamte «visual information of humanity» beinhalten.[39] Demgegenüber unterstreichen Positionen aus den Critical Dataset Studies, dass Bilddatenbanken wie LAION keineswegs vollständige Repräsentationen visueller Kultur darstellen und solche Selbstdarstellungen vielmehr als ideologische Legitimationen gegenwärtiger Machine-Learning-Projekte verstanden werden müssen:

> This shift […], we argue, represents a deeper political shift towards an understanding of large visual models as ‹complete› models of visual culture. […] The popularization of the term ‹foundation model› by researchers at Stanford and elsewhere suggests this understanding indeed also serves as the ideological foundation of contemporary machine learning research.[40]

Die Behauptung von Stability AI, die Software würde die ganze Menschheit visuell repräsentieren, reiht sich in eine Reihe von Selbstdarstellungen aus der Machine-Learning-Industrie ein,[41] die mit ihren Ansprüchen auf Universalität und Totalität nicht zuletzt ihre eigene Relevanz zu belegen versuchen.

«Poor images» meet Plattformrealismus

Trotz der starken Verzerrungen und Zurichtungen des LAION-Datensatzes entspricht die Visualität eines LAION-Bildes doch am ehesten dem, was Hito Steyerl 2009 als «poor image» bezeichnet hat: Ein durchschnittliches Internetbild mit schlechter Auflösung und kaum ästhetischem Anspruch, «distributed for free, squeezed through slow digital connections, compressed, reproduced, ripped, remixed, as well as copied and pasted into other channels of distribution»,[42] irgendwann zwischen 2008 und 2022 von irgendwem irgendwo hochgeladen, mit einer obskuren Caption und unklarem Copyright versehen, dann heruntergeladen, herunterskaliert und umbenannt für die Datenbank. Das wird mir klar, als ich auf eine Reihe von Bildern aus LAION-2B stoße, die für ein Pilotprojekt automatisiert neu untertitelt wurden. Ich scrolle durch eine PDF-Datei mit Bildern und Captions und bleibe an einem Bild mit der Caption «A man flying a kite on a beach» hängen.

37 Vgl. ebd, 2 f.

38 Vgl. Christo Buschek, Jer Thorp: Models All the Way Down [Beitrag auf Projektseite], *Knowing Machines*, 26.3.2024, *knowingmachines.org/models-all-the-way* (11.3.2025).

39 Vgl. Stable Diffusion Public Release [Pressemitteilung], *Stability AI*, 22.8.2022, *stability.ai/news/stable-diffusion-public-release* (18.12.2025).

40 Fabian Offert, Thao Phan: A Sign That Spells. DALL-E 2, Invisual Images and The Racial Politics of Feature Space, *arXiv*, 26.10.2022, 1–4, hier 2, *doi.org/10.48550/arXiv.2211.06323*.

41 Vgl. Lisa Rein, Sabine Wirth: Exzesse des Indizierens. Von fotografischen Archiven zur Bildklassifizierung im Plattformkapitalismus, in: *Fotogeschichte. Beiträge zur Geschichte und Ästhetik der Fotografie*, Jg. 45, Nr. 177, 2025, 37–45, hier 42.

42 Vgl. Hito Steyerl: In Defense of the Poor Image, in: *e-flux Journal*, Nr. 10, 2009, 1–9, hier 1.

Es ist ein Foto im Querformat, in schlechter Auflösung, geschossen am Strand in Richtung Meer, der Horizont verläuft diagonal von links unten nach rechts oben, sodass das ganze Bild in Schräglage ist. Eine Person ist zu sehen, mittelalt, korpulent, von mir als Mann gelesen, mit hellrosa Haut, in beigen Caprishorts und beigem Anglerhut, das Gesicht vom Schatten des Huts und einer kleinen Sonnenbrille verdeckt. Aufgrund des schrägen Horizonts rutscht er fast aus dem Bild heraus. Mit beiden Händen hält er eine Drachenschnur, der kleine, bunte Drachen liegt rechts im Bild am Boden im Sand, anscheinend kein Glück mit dem Wind. Im Vordergrund des Bildes, nah an der Kamera, leicht unscharf, ist so etwas wie das Gestänge eines Campingstuhls sichtbar, es wirkt nicht gewollt dort platziert, eher versehentlich ins Bild geraten. Das eigentliche Zentrum des Bildes ist leer. Der Himmel ist bedeckt. Die Farben des Bildes sind gräulich, das Farbspektrum begrenzt; es sieht sehr nach einem privaten Schnappschuss aus, vermutlich entweder mit einer Digitalkamera oder einem Smartphone geschossen, aus der Perspektive einer auf dem Boden sitzenden Person.

Wenn wir nun die automatisch generierte Caption des Bildes, «A man flying a kite on a beach», 2025 als Prompt für den auf LAION trainierten Text-to-Image-Generator Stable Diffusion verwenden, erhalten wir ein etwas anderes Bild (vgl. Abb. 2). Es wirkt wie eine optimierte Version jenes *poor image*: Im Zentrum des quadratischen Bildes ist eine muskulöse, braun gebrannte, von mir ebenfalls männlich gelesene Person abgebildet, die energisch einen bunten Drachen lenkt, im Hintergrund verläuft die gerade Horizontlinie des Wassers, die Sonne steht hoch und kreiert eine atmosphärische Lichtstimmung mit starken Kontrasten; die Farben leuchten, die Sättigung ist hoch; die ganze Ästhetik ist *glossy* und irgendwie hyperrealistisch. Auf den ersten Blick sieht es aus wie ein Foto, auf den zweiten eher wie ein Bild, das Fotografie als Stil imitiert, ein Stockfoto oder eine fotorealistische Malerei. Roland Meyer bezeichnet diese Ästhetik KI-generierter Bilder als «Plattformrealismus»:[43] In ihnen hallen stereotype Vorstellungen fotografischer Ästhetik und Bildtechniken wider, die in smartphonebasierten Plattformumgebungen beliebt sind, wie z. B. bestimmte digitale Filter, die auf Social-Media-Plattformen Konjunktur haben. Letztlich könnte man sagen, dass jedem Prompt ein unsichtbares «in the style of HD digital photography» hinzugefügt wird und es sich, wie Meyer schreibt, beim Ergebnis nicht um «Bilder der Realität, sondern [um] Bilder von Bildern» handelt.[44] Wie diese spezifische Ästhetik zustande kommt, obwohl LAION zu großen Teilen aus den *poor images* des World Wide Webs der 2000er und 2010er Jahre, aus PowerPoint-Slides, Thumbnails und Produktfotos von Onlineshops besteht, wird im folgenden Abschnitt deutlich.

Tatsächlich basiert die aktuelle Version von Stable Diffusion auf einer kleineren, stärker kuratierten Version von LAION, nämlich aus einem speziell für Stable Diffusion aufgebauten Datensatz namens LAION-Aesthetics.[45] Für

43 Roland Meyer: «Platform Realism». AI Image Synthesis and the Rise of Generic Visual Content, in: *Transbordeur*, Nr. 9, 2025, 1–18, hier 11, doi.org/10.4000/13dwq.

44 Ders.: Bilder von Bildern. Praktiken und Prozesse der KI-Bildgenerierung, in: *Fotogeschichte. Beiträge zur Geschichte und Ästhetik der Fotografie*, Jg. 45, Nr. 177, 2025, 30–36, hier 36.

45 Vgl. Christoph Schuhmann: LAION-Aesthetics [Beitag auf Unternehmensblog], LAION, 16.8.2022, *laion.ai/blog/laion-aesthetics* (26.9.2025).

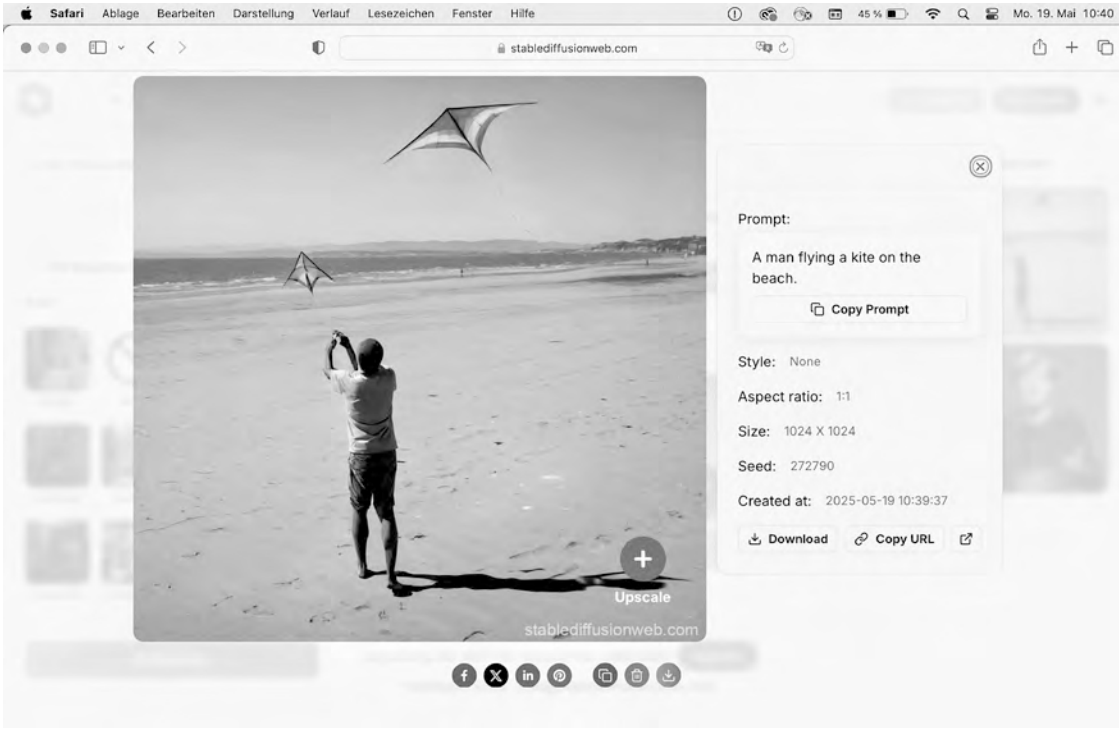

Abb. 2 **Mit Stable Diffusion generiertes Bild zum Prompt «A man flying a kite on a beach»**

LAION-Aesthetics wurde der gesamte LAION-Datensatz anhand einer Skala von 0 (nicht ästhetisch) bis 10 (ästhetisch) durch ein algorithmisches Modell in einzelne Klassen sortiert. Das Ergebnis hat LAION auf einer Webseite veröffentlicht (vgl. Abb. 3).[46] Gleich auf den ersten Blick wird deutlich, dass im oberen Bereich, mit Werten ab 5 oder 6 aufwärts, vor allem hyperrealistische computergenerierte Bilder, die wie gemalt aussehen, angesiedelt sind. Die Sättigung ist hoch, die Farbtöne hell, als Motive dienen hauptsächlich normschöne Körper (sprich: weiße, schlanke Frauen) und Landschaften.[47] Am unteren Ende der Skala dagegen befindet sich der Rest der Bilder, die *poor images*: verwackelte Bilder mit niedriger Auflösung, Screenshots mit unleserlichen Texten, unscharfe Bilder von Buchcovern, Platzhalterbilder mit Wasserzeichen – «the trash that washes up on the digital economies' shores», wie Steyerl schreibt.[48] Zum Trainieren von Stable Diffusion (und im Übrigen auch für das Finetuning der Konkurrenz-Software Midjourney[49]) wurden nur die Bilder verwendet, die auf dieser Ästhetik-Skala mindestens einen Wert von 5 erzielt haben.[50]

Um zu verstehen, auf welcher Basis das für LAION-Aesthetics eingesetzte Modell seine Zuordnungen vornimmt, müssen wir allerdings noch eine Schicht tiefer ansetzen. Das Modell beruht hauptsächlich auf zwei Datenbanken aus dem Bereich des *automatic aesthetic assessment*, die jeweils aus hunderttausenden Bildern bestehen und deren zweite Dateneinheit jeweils ein numerischer Wert

46 Vgl. Christoph Schuhmann: Aesthetic Subsets in LAION 2170337258 Samples [Demo-Webseite], *Christoph-Schuhmann.de*, o. D., *captions.christoph-schuhmann.de/aesthetic_viz_laion_sac+logos+ava1-l14-linearMSE-en-2.37B.html* (2.12.2025)

47 Vgl. Impett: Computation and Beauty.

48 Steyerl: In Defense of the Poor Image, 1.

49 Vgl. Buschek, Thorp: Models all the Way Down.

50 Vgl. Schuhmann: LAION-Aesthetics.

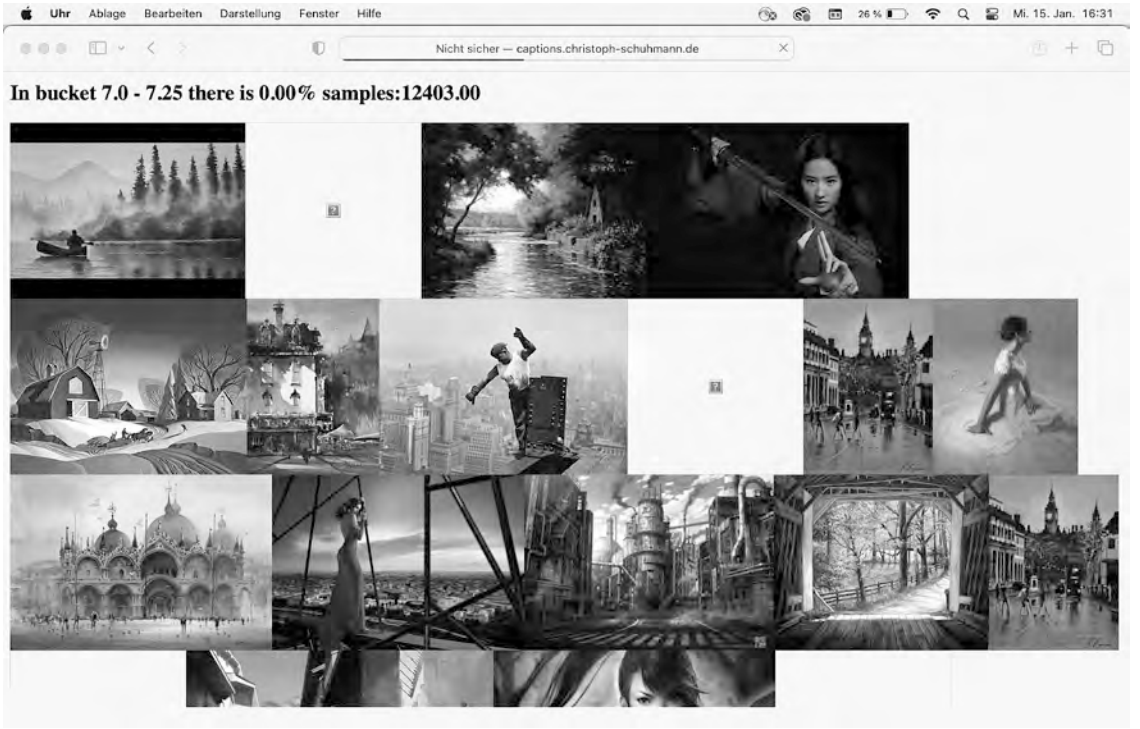

Abb. 3 **Beispielbilder aus dem Datensatz LAION-Aesthetics mit Wert 7–7,25 (von der Demo-Website von Christoph Schuhmann, 2022)**

zwischen 0 bzw. 1 und 10 bildet:[51] zum einen auf AVA, einer bereits seit 2012 existierenden Open-Source-Datenbank, und zum anderen auf SAC, einer im LAION-Umfeld selbst entwickelten Datenbank, die sich freiwillige Communityarbeit zunutze macht.

AVA: von der Foto-Sharing-Community zur Datenbank

Die Datenbank AVA (kurz für Aesthetic Visual Analysis)[52] wurde 2012 veröffentlicht und gilt seitdem als Goldstandard und als am häufigsten verwendete Datenbank im Bereich des *automatic aesthetic assessment*.[53] AVA enthält ca. 250.000 Bilder, die allesamt der Webseite DPChallenge.com entnommen wurden – einer Seite, die seit 2002 eine Amateur*innen-Foto-Community beheimatet und den frühen Foto-Sharing-Seiten der 2000er Jahre zuzurechnen ist.[54] Das «DP» im Titel der Webseite steht für *digital photography*, und «challenge» verweist darauf, dass auf der Webseite Fotografien von Nutzer*innen in Wettbewerben zu vorgegebenen Themen gegeneinander antreten und dann von der Community aller Nutzer*innen bewertet und gerankt werden, woraufhin Gewinnerbilder gekürt werden. In den Augen der AVA-Entwickler*innen ist der auf DPChallenge angesammelte Fundus von Bildern und numerischen Bewertungen eine Goldgrube:

51 Ebd.

52 Vgl. Naila Murray, Luca Marchesotti, Florent Perronnin: AVA. A Large-Scale Database for Aesthetic Visual Analysis, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2012, 2408–2415, doi.org/10.1109/CVPR.2012.6247954.

53 Vgl. Matteo Bodini: Will the Machine Like Your Image? Automatic Assessment of Beauty in Images with Machine Learning Techniques, in: *Inventions*, Bd. 4, Nr. 3, Sep. 2019, 1–18, hier 10, doi.org/10.3390/inventions4030034.

54 Vgl. Katrina Sluis: Photography Must Be Curated! Part Four: Survival of the Fittest Image [Essay], *Fotomuseum Winterthur*, 23.1.2020, fotomuseum.ch/de/2020/01/23/survival-of-the-fittest-image (29.1.2025).

> The number of votes per image ranges from 78 to 549, with an average of 210 votes. Such score distributions represent a gold mine of aesthetic judgments generated by hundreds of amateur and professional photographers with a practiced eye. We believe that such annotations have a high intrinsic value because they capture the way hobbyists and professionals understand visual aesthetics.[55]

Erstens stellen aus informatischer Sicht die hohe Anzahl an Bewertungen pro Bild einen enormen Mehrwert dar, zweitens stammen die Bewertungen den AVA-Entwickler*innen zufolge von einer Community von professionellen und Amateur-Fotograf*innen, deren Bildbewertungen aufgrund ihrer Expertisen als wertvoll eingeschätzt werden. Drittens werden die Bilder nicht in Form von Freitext oder binären ‹Like›/‹Don't Like›-Interaktionen bewertet, sondern auf einer numerischen Skala – damit bilden sie die optimale Voraussetzung für eine statistische Weiterverarbeitung.

Dabei bleibt von den AVA-Entwickler*innen unkommentiert, dass die verwendeten Daten in vielerlei Hinsicht von Einschränkungen und Besonderheiten markiert sind: Zunächst einmal ist der Datensatz von einem «temporal bias»[56] gezeichnet, also von starker zeitlicher Begrenztheit, die daraus resultiert, dass die verwendeten Daten nur einen Zeitraum von rund zehn Jahren abbilden, von der Veröffentlichung der DPChallenge-Webseite im Jahr 2002 bis zur Entwicklung von AVA 2012. Folglich wurden fast alle auf der Webseite hochgeladenen Fotografien mit digitalen (Spiegelreflex-)Kameras aufgenommen und sind von den medientechnischen Bedingungen und visuellen Paradigmen der 2000er Jahre geprägt. Darüber hinaus spiegelt AVA den Geschmack einer spezifischen Gruppe von semiprofessionellen, netzaffinen Foto-Enthusiast*innen wider.[57] Demografisch konstituiert sich diese Gruppe hauptsächlich aus mittelalten nordamerikanischen Männern,[58] von denen sich viele in ihren Profilen als ‹*techies*› outen. Wie in Huxleys dystopischem Roman wird das AVA-Universum von Männern und ihren Blicken auf die Welt dominiert.[59] Besonders deutlich wird das, wenn man der nebensächlichen Feststellung der AVA-Entwickler*innen, wonach diejenigen Foto-Challenges mit den *meisten* Bewertungen pro Bild «nude subjects or lingerie» thematisieren,[60] genauer nachgeht. Unter den Challenges, die Eingang in das AVA-Datenset gefunden haben, sind die mit den meisten Bewertungen pro Bild die Challenges «Nude III», «Nude IV», «Nude V» und «Lingerie». Fast alle Gewinnerfotos dieser Challenges zeigen normschöne, als weiblich inszenierte Körper in sexualisierten Posen und entsprechen in Bildsprache und -sujet geradezu klischeehaft *male-gaze*-Fotografien.[61] Generell sind die Gewinnerfotos auf DPChallenge von starker Sättigung, hohem Kontrast und atmosphärischem Licht geprägt. Es sind viele Schwarzweißbilder darunter, und überproportional häufig zeigen sie Landschaften oder Stillleben, oftmals mit Spiegelungen oder Lichtspielereien, Menschen sind selten zu sehen (vgl. Abb. 4).[62] Zu dieser Erkenntnis komme ich nicht nur über meine Bildschirmspaziergänge durch die Tiefen von DPChallenge[63] und dank der

55 Murray, Marchesotti, Perronnin: AVA, 2409.

56 Thomas Smits, Melvin Wevers: The Agency of Computer Vision Models as Optical Instruments, in: *Visual Communication*, Bd. 21, Nr. 2, Mai 2022, 329–349, hier 340, doi.org/10.1177/14703572211992097.

57 Vgl. Bodini: Will the Machine Like Your Image?, 13.

58 Vgl. Impett: Computation and Beauty.

59 Vgl. Divya Khasa: Resistance and Representation. Women, Technology, and Power in *Brave New World* and *Orphan Black*, in: *The Text*, Bd. 6, Nr. 2, 2024, 47–61, hier 54.

60 Murray, Marchesotti, Perronnin: AVA, 6.

61 Vgl. o. A.: Challenge History, sortiert nach «Average Votes», *DPChallenge*, dpchallenge.com/challenge_history.php?order_by=7d&page=1 (3.12.2025).

62 Vgl. Impett: Computation and Beauty.

63 Im Gegensatz zu vielen anderen Machine-Learning-Ressourcen ist die Webseite glücklicherweise nach wie vor online und wird anscheinend sorgfältig gepflegt.

Abb. 4 **Beispielbilder aus AVA, mit «landscape» getaggt und mit durchschnittlicher Bewertung in Klammern**

Analysen des Digital-Humanities-Forschers Leonardo Impett, sondern auch anhand der von den AVA-Entwickler*innen eigenhändig vorgenommenen thematischen Klassifizierung der Bilder. Dieser zufolge sind die fünf häufigsten *semantic tags* ‹Nature›, ‹Black and White›, ‹Landscape›, ‹Still Life› und ‹Macro›.[64]

Die normierenden Tendenzen, die sich daraus ergeben, dass als Grundlage für die Bildbewertungen die äußerst spezifischen Geschmacksurteile einer relativ homogenen demografischen Gruppe herangezogen wurden, werden im AVA-Datensatz noch verstärkt durch die Verwendung klassischer Methoden aus dem Bereich des Maschinellen Lernens: Um stabile Ergebnisse rund um einen Mittelwert zu erzielen, werden für AVA alle Bilder aussortiert, bei denen die DPChallenge-Votes sehr weit auseinanderliegen, bei denen es also geringe Einigkeit in der Beurteilung gibt. «Images with a high variance seem more likely to be edgy or subject to interpretation, while images with a low variance tend to use conventional styles or depict conventional subject matter», konstatieren die AVA-Entwickler*innen und begründen damit die Entfernung von «ambiguous images» aus dem Datensatz.[65] Dieser Normalisierungseffekt ist typisch für statistische Machine-Learning-Verfahren, die mithilfe von Gauß'scher Normalverteilung Wahrscheinlichkeiten berechnen. «DNNs [*deep neural networks*] prefer *normal* images», resümiert der Informatiker Matteo Bodini und weist darauf hin, dass dieser Hang zum Mittelmaß nicht unbedingt den Meinungen von Ästhetik-Expert*innen entspricht, welche die AVA-Macher*innen vorgeben abzubilden.[66] Seine an die eigene Disziplin gerichtete Frage «Which beauty? Which expert?»[67] bleibt von AVA unbeantwortet und verschwindet hinter mathematisch überzeugenderen Argumenten rund um Quantität und Mittelwert.[68] Dass es dieser an repräsentativen Durchschnittswerten interessierten Logik eigentlich widerspricht, ein und dieselbe Gruppe von Personen für die Herstellung der Bilder *und* ihre Bewertung heranzuziehen, wird von den AVA-Entwickler*innen nicht thematisiert und muss wohl auf informatischen Pragmatismus zurückgeführt werden.

64 Vgl. Murray, Marchesotti, Perronnin: AVA, 3.

65 Vgl. ebd., 6 f.

66 Bodini: Will the Machine Like Your Image?, 12, Anm. LR.

67 Ebd.

68 Vgl. dazu auch Hito Steyerl: Mean Images, in: *New Left Review*, Nr. 140/141, 2023, 82–97.

SAC oder «Wie gut gefällt dir dieses Bild?»

Auch SAC (Simulacra Aesthetic Captions), die zweite für LAION-Aesthetics verwendete Datenbank, beinhaltet knapp 250.000 Bilder.[69] Im Gegensatz zu AVA wurde SAC allerdings 2022 speziell für die Optimierung von Bilddatenbanken wie LAION und deren Implementierungen angelegt, initiiert von zwei an LAION beteiligten Entwickler*innen, Katherine Crowson und John David Pressman. Die an der *ground truth* von SAC beteiligten menschlichen Betrachter*innen trugen freiwillig mit ihrer Bewertungsarbeit zur Erstellung des Datensatzes bei. SAC ist mit Unterstützung von Nutzer*innen aus den zwei Discord-Channels zu Stable Diffusion und Glide (einem weiteren Text-to-Image-Generator) entstanden. Grundlage der Datenbank sind hier keine fotografischen Bilder, sondern mit Software generierte Bilder, die von den Discord-Nutzer*innen mithilfe von Stable Diffusion, Glide und weiteren Text-to-Image-Generatoren erstellt und anschließend bewertet wurden. Die Nutzer*innen der beiden Channels wurden von Crowson und Pressman zunächst dazu aufgefordert, mit eigenen Prompts Bilder zu generieren, diese im Channel zu teilen und anschließend alle so erstellten Bilder anhand der Frage «Wie gut gefällt dir dieses Bild?»[70] auf einer Skala von 1 bis 10 zu bewerten.[71]

Der medientechnologische und ästhetische *temporal bias* von SAC ist also noch stärker ersichtlich als der von AVA, handelt es sich doch ausschließlich um Bilder, die in den Jahren 2021 und 2022 produziert wurden und die von der oben beschriebenen spezifischen Ästhetik der Text-to-Image-Generatoren gezeichnet sind. Es mutet zunächst absurd an, dass Stable Diffusion (als Text-to-Image-Generator, der auf LAION trainiert wurde) eingesetzt wird, um Bilder zu generieren, die dann wiederum dazu beitragen, eine optimierte Version von LAION (und damit auch von Stable Diffusion) zu kreieren. Eine solche selbstreferenzielle Vorgehensweise ist allerdings im Zeitalter von «crawl over curate» charakteristisch für Datenbanken und -modelle. Die daraus resultierenden «stochastic parrots»[72] – auf generative Bilderzeugung bezogen auch schon als «Habsburg AI» bezeichnet[73] – verstärken die normativen Tendenzen großer Modelle, indem sie, wie im Fall von SAC, den Möglichkeitsraum potenzieller Bilder und Visualitäten weiter einengen.

Auch die genaue Untersuchung der an der Erstellung von SAC beteiligten Nutzer*innengruppe ergibt im Vergleich zu AVA sowohl quantitativ als auch qualitativ noch stärkere Homogenität, auf welche die beiden Entwickler*innen selbst hinweisen: Sie räumen in der Beschreibung des Datensatzes ein, dass sich an der Bildgenerierung und -bewertung nur «eine Handvoll» der insgesamt 400 Channel-Nutzer*innen beteiligen, deren ästhetische Vorlieben in der Folge den gesamten Datensatz dominieren. Auch stellen sie unmissverständlich klar, dass es sich ihrer Ansicht nach bei dieser Nutzer*innengruppe größtenteils um «power users and developers of AI art» handelt. Diese seien als

69 Vgl. John David Pressman, Katherine Crowson, Simulacra Captions Contributors: Simulacra Aesthetic Captions, GitHub, 4.7.2022, github.com/JD-P/simulacra-aesthetic-captions (15.1.2025).

70 Vgl. Francis Hunger im Gespräch mit Alexa Steinbrück: Repräsentationsweisen Künstlicher Intelligenz und wie eine künstlerische Lehre aussehen kann, in: Inke Arns u. a. (Hg.): *Training the Archive* [Ausstellungsmagazin zum gleichnamigen Forschungsprojekt 2020–2023], Köln 2024, 129–136, hier 135.

71 Vgl. Pressman, Crowson, Simulacra Captions Contributors: Simulacra Aesthetic Captions.

72 Emily M. Bender u. a.: On the Dangers of Stochastic Parrots. Can Language Models Be Too Big? 🦜, in: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 1.3.2021, 610–623, dl.acm.org/doi/10.1145/3442188.3445922.

73 Jathan Sadowski @jathansadowski: I coined a term …, X, 13.2.2023, x.com/jathansadowski/status/1625245803211272194 (12.3.2025).

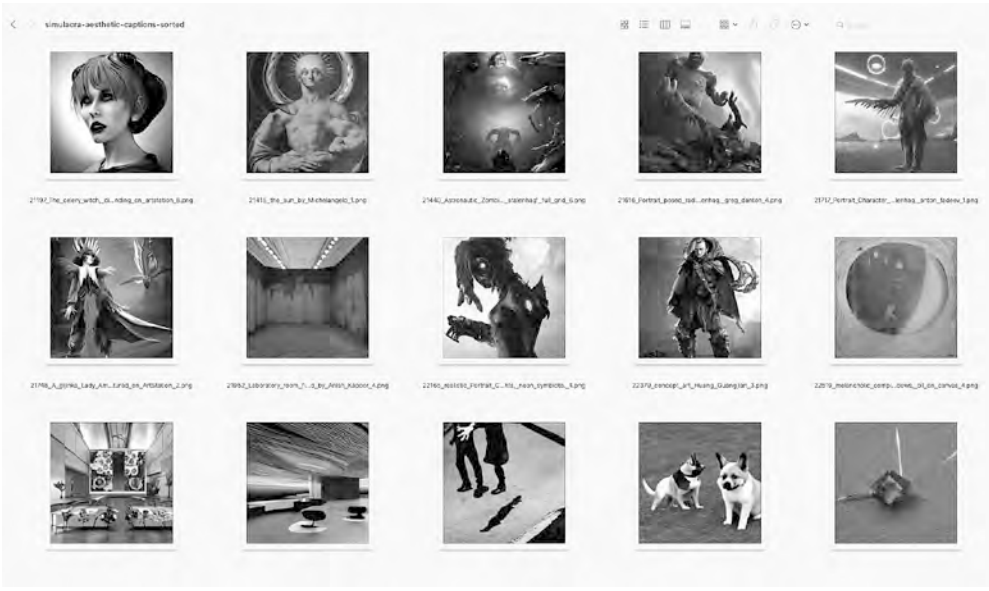

Abb. 5 Ausschnitt aus der Datenbank SAC

«WEIRD», d. h. «Western, Educated, Industrialized, Rich, and Democratic» zu kategorisieren, woraus die Autor*innen ableiten, «that their aesthetic feedback is going to lean [sic] STEM [Science, Technology, Engineering, Mathematics], fantasy, nerdy, esoteric, etc».[74] Diese bemerkenswerte Schlussfolgerung bestätigt sich auch auf stichprobenhaften Streifzügen durch den Datensatz: Die Prompts lesen sich zu großen Teilen wie Sci-Fi-Romantitel, gemischt mit dystopischen Fantasy-Elementen und naturwissenschaftlichen Referenzen, wie z. B. «the machine_elves_dance_around_my_mind» und «unreal_engine_render_of_the_biohazard_radiation_protection_suit». Aber auch popkulturelle Referenzen und bekannte Persönlichkeiten (hauptsächlich Männer) wie Barack Obama, Mark Zuckerberg, Drake, John von Neumann, Karl Marx, Charles Manson und Elon Musk sind vertreten, wie z. B. im Prompt «12811_Portrait_of_elon_musk_with_cool_glasses_in_witcher_trending_4k». Viele Bilder zeigen psychedelische Motive, Fabelwesen, immer wieder Engel und Todessymboliken, fremde Galaxien und fiktive Technologien. Die Farben sind schrill, die Formen verflüssigt, die typische Glitch-Ästhetik der frühen Text-to-Image-Generatoren zieht sich durch. Für mich sieht es aus, als ob hier Silicon-Valley-Ideologie auf Nerdkultur und Amateur*innenkunst trifft.

Wie schwierig es ist, selbst innerhalb einer kleinen, homogenen Gruppe konsistente und verlässliche Werte hinsichtlich der ästhetischen Bewertung von Bildern zu erzielen, wird aus der Projektbeschreibung deutlich: Crowson und Pressman sahen sich nach ersten Tests gezwungen, eine Art Eignungstest durchzuführen, um Nutzer*innen auf ihre Biases hin abzuklopfen. Die Entwickler*innen beschreiben diese Notwendigkeit folgendermaßen:

74 Pressman, Crowson, Simulacra Captions Contributors: Simulacra Aesthetic Captions, Anm. LR.

> [D]uring data collection it was observed that some users would downrate objectively well drawn pictures because they're ‹scary›, or included public figures they didn't like such as the US president […]. Rather than try to force people to go against their intuition, which seemed somewhere between tyrannical and futile, an aesthetic survey was implemented on signup that records user's behavior around these confounding factors in aesthetic judgment.

Abb. 6 **Bilder aus SAC, anhand derer Befangenheiten von Bild-Bewerter*innen ausfindig gemacht werden sollen**

Vor der Freigabe des Bild-Bewertungs-Interfaces mussten alle Nutzer*innen ein solches «aesthetic survey» absolvieren, um ihre jeweiligen Befangenheiten und Neigungen identifizieren zu können und ihre Fähigkeiten, Bilder ‹neutral› zu bewerten, zu überprüfen. Die ‹Fehler›-Kategorien, die mithilfe von 20 Beispielbildern identifiziert werden sollten (vgl. Abb. 6), beziehen sich zum einen auf bildtechnische Fragen, wie Wasserzeichen im Bild – hier wäre aus Crowsons und Pressmans Sicht *richtig* gewesen, wenn Nutzer*innen aufgrund des Wasserzeichens das Bild schlecht bewertet hätten. Zum anderen betreffen sie aber auch die Bildsujets, wie beispielsweise bei der Kategorie «weak prompt fit», bei der aus Sicht der Entwickler*innen eine geringe Übereinstimmung zwischen Prompt und Bild vorliegt und die Bewertung dementsprechend niedrig hätte ausfallen müssen. Oder bei der Kategorie «Quality vs. Subject», bei der das Test-Bild wie ein Ölgemälde von Adolf Hitler aussieht und offenbar sichergestellt werden soll, dass Nutzer*innen sich in der Bewertung der «Qualität» nicht negativ (oder positiv?) vom Bildsujet beeinflussen lassen sollen. Konkrete Bildsujets, bei denen aus Sicht von Crowson und Pressman die Gefahr besteht, dass Nutzer*innen sich dazu verleiten lassen könnten, «objectively well drawn pictures» schlecht zu bewerten, umfassen zudem die Kategorien «Ethnic art/Racism», «Islam Sentiment», «Fantasy/Witch», «Cute/Feminine» und «Heroic/Masculine».

Zunächst illustriert diese Auflistung in Kombination mit den jeweiligen Beispielbildern die von den beiden Entwickler*innen identifizierten visuellen Präferenzen und die potenziellen Biases der Bildbewerter*innen, die sich offensichtlich auch in rassistischen, misogynen oder islamfeindlichen Ressentiments und dem Abarbeiten an Männlichkeitsbildern äußern. Darüber hinaus wird an diesem Ringen um vermeintlich neutrale Bildbewertungen aber vor allem deutlich, wie wenig die Frage «Wie gut gefällt dir dieses Bild?» zu objektiven, vom Bildsujet unabhängigen ästhetischen Einschätzungen führt. Im Gegenteil ist

diese Frage von erstaunlicher Offenheit, wie die Künstlerin und Programmiererin Alexa Steinbrück betont, die sich mit der Datenbank eingehend auseinandergesetzt hat:

> ‹Wie gut gefällt dir dieses Bild?› [kann] natürlich alles Mögliche bedeuten […]. Das kann meinen: ‹Mir gefällt der Inhalt nicht›, ‹Mir gefällt die Belichtung nicht›, ‹Mir gefällt die Auflösung nicht› oder ‹Mir gefällt die JPEG-Kompression nicht›. Das ‹Gefallen› lässt so viel offen und trotzdem ist es jetzt ein zentraler Bestandteil der Modelle, mit denen wir arbeiten.[75]

Dass diese Modelle unter informatischen Fachtermini wie *automatic aesthetic assessment* verhandelt werden, verschleiert, dass Datenbanken wie AVA, SAC und LAION-Aesthetics weniger formale ästhetische Beurteilungen versammeln als vielmehr spontane Geschmacksäußerungen einer überschaubaren und wenig diversen Gruppe von Menschen.

Die Quantifizierung von Geschmack

Geschmack ist, wie medienwissenschaftliche Positionen im Anschluss an Pierre Bourdieus diesbezügliche Thesen argumentieren, einerseits zentrales Element sozialer Distinktion und Ausdruck von Klassenzugehörigkeit, andererseits aber auch eine transformative Variable, die in einem ständigen Aushandlungsprozess immer aufs Neue performativ wird.[76] An diesem Prozess sind nicht nur die soziale Position und Prägung eines Individuums beteiligt. In digitalen Plattformumgebungen mischen sich vielmehr auch medientechnologische Infrastrukturen, algorithmische Ordnungen und kapitalistische Logiken aktiv in die Aushandlungen von Geschmack, in das *taste making* ein.[77] Wenn mit Johannes Paßmann und Cornelius Schubert *taste making* als konstitutives Element von Social-Media-Plattformen verstanden werden kann (als Praktiken ästhetischer Urteile, die wiederum soziale Distinktion ermöglichen), so zeigt sich hier, wer/was an diesen Verfahren (unbemerkt) beteiligt ist:[78] Bilddatenbanken wie AVA, SAC und LAION-Aesthetics sind machtvolle, wenn auch für Nutzer*innen kaum sichtbare Akteure für dieses *taste making* in plattformkapitalistischen Umgebungen. Dabei ermöglichen sie es, Geschmack anhand von Ratings zu quantifizieren, und bedienen somit eine grundsätzliche Anforderung von Plattformen: popkulturelle Praktiken in messbare Listen- und Rankingsysteme zu übertragen.[79] Durch algorithmisch kuratierte Feeds und Empfehlungssysteme strukturieren Plattformen Sichtbarkeiten anhand von messbaren und sortierbaren numerischen Systemen, stets mit dem Ziel, User*innen-Engagements zu optimieren und maximieren.

Die genaue Betrachtung der Bilddatenbanken aus dem Bereich des *automatic aesthetic assessment* macht deutlich, dass *taste making* nicht nur auf den sichtbaren *surfaces* der Plattformen stattfindet, sondern weit in den *subfaces* von Plattformumgebungen verankert ist. Dabei entstehen rekurrierende Feedbackschleifen, in denen die Quantifizierung von Geschmack einerseits durch Verwertungs-

75 Hunger, Steinbrück: Repräsentationsweisen Künstlicher Intelligenz, 136.

76 Vgl. Johannes Paßmann, Cornelius Schubert: Kritik der digitalen Urteilskraft. Soziale Praktiken der Geschmacksbildung im Internet, in: *Mittelweg 36*, Jg. 30, Nr. 1, 2021, 60–84, hier 61.

77 Vgl. Niko Pajkovic: Algorithms and Taste-Making. Exposing the Netflix Recommender System's Operational Logics, in: *Convergence. The International Journal of Research into New Media Technologies*, Bd. 28, Nr. 1, Feb. 2022, 214–235, hier 230, doi.org/10.1177/13548565211014464.

78 Vgl. Paßmann, Schubert: Kritik der digitalen Urteilskraft, 61 f.

79 Vgl. Ralf Adelmann: *Listen und Rankings. Über Taxonomien des Populären*, Bielefeld 2021.

logiken von Plattformen notwendig gemacht wird und andererseits überhaupt erst durch vorausgehende Nutzer*innenpraktiken, wie diejenigen der DPChallenge-Community, möglich gemacht werden. Ausgehend vom eingangs entworfenen Bild des Archipel ließe sich sagen, dass in der Übergangszone zwischen Plattform und Datenbank Nutzer*innenpraktiken, Plattforminteressen, ökonomische Verwertungslogiken und informatische Datafizierungen aufeinandertreffen und immer wieder von Neuem Einfluss aufeinander nehmen. Das zeigt sich auch darin, dass die von Social-Media-Plattformen perfektionierte ‹Wie gefällt dir das?›-Logik Eingang in den Bereich des *automatic aesthetic assessment* gefunden hat. Anfängliche Versuche, formal-ästhetische Kriterien wie Symmetrien, Bildfarben und -auflösung oder Erkenntnisse aus dem Bereich der «neuro-aesthetics» zu übernehmen, sind in den Hintergrund geraten.[80] Gleichzeitig findet hier auch eine gewaltige Übersetzungsbewegung statt, in der «eine (imaginäre) ästhetische Kategorie in sichtbare und zählbare Einheiten» umgewandelt werden muss.[81] Diese Bewegung stellt sich, ähnlich wie das brachiale Ringen zwischen Wasser und Land, das die «shifting landscape» des Archipels prägt,[82] als andauernde, machtvolle Aushandlung dar. Darum komme ich im letzten Schritt noch einmal auf die bildpolitischen Konsequenzen zurück, die diese mächtigen Interface-Prozesse für die auf den *surfaces* verhandelten Fragen nach Sichtbarkeit, Repräsentation und Subjektivierung haben.

«Regimes of recognition»

Der etwas hilflos wirkende Versuch der SAC-Entwickler*innen, die Blicke ihrer Kollaborateur*innen zu disziplinieren, demonstriert, wie schlecht sich visuelle Wahrnehmung und Geschmack formalisieren lassen – und wie sehr sie von historischen, politischen, kulturellen und medientechnologischen Kontexten abhängen. Die Visual Culture Studies haben dies mit Verweis unter anderem auf Walter Benjamins Thesen zur Historizität der Wahrnehmung vielfach betont.[83] Zahlreiche kritische Abhandlungen zum *gaze* (insbesondere aus der feministischen Theorie) haben aufgezeigt, dass Blicke und die Bewertung von Bildern immer in Machtgefüge verwickelt sind.[84] In Hinblick auf den von Plattforminteressen geprägten und von algorithmischen Infrastrukturen implementierten ‹Blick› sind Begriffe wie «algorithmic gaze» oder «data gaze» eingeführt worden, um die spezifischen Machthierarchien und Subjektivierungsprogramme zu erfassen, die mit datengetriebenen, plattformkapitalistischen Blickregimen einhergehen.[85] Louise Amoore begreift sie als «regimes of recognition», die aktiv an der Herstellung von «recognizability» mitwirken:

> Machine learning algorithms do not merely recognize people and things in the sense of identifying faces, threats, vehicles, animals, languages – they actively generate recognizability as such, so that they decide what or who is recognizable as a target of interest in an occluded landscape.[86]

80 Vgl. Bodini: Will the Machine Like Your Image?, 4, 8.

81 Paßmann, Schubert: Kritik der digitalen Urteilskraft, 78.

82 Snelting: Other Geometries.

83 Vgl. Marius Rimmele, Bernd Stiegler: *Visuelle Kulturen/Visual Culture zur Einführung*, Hamburg 2012, 65 f.

84 Vgl. u. a. Laura Mulvey: Visual Pleasure and Narrative Cinema, in: *Screen*, Bd. 16, Nr. 3, Herbst 1975, 6–18, doi.org/10.1093/screen/16.3.6.

85 Dan M. Kotliar: Data Orientalism. On the Algorithmic Construction of the Non-Western Other, in: *Theory and Society*, Bd. 49, Nr. 5/6: *The Sociology of Digital Technology. A Special Issue*, Nov. 2020, 919–939, jstor.org/stable/48735077; David Beer: *The Data Gaze. Capitalism, Power and Perception*, London 2019, doi.org/10.4135/9781526463210.

86 Louise Amoore: *Cloud Ethics. Algorithms and the Attributes of Ourselves and Others*, Durham, London 2020, 69.

Zahlreiche Beispiele haben in den letzten Jahren gezeigt, dass sich diese soziotechnischen *regimes of recognition* zunehmend auch auf Social-Media-Plattformen manifestieren, die ja schon lange zentrale Aushandlungsorte für Kämpfe um Sichtbarkeit bilden. Plattformbetreiber*innen treten als machtvolle Akteur*innen in diesen Aushandlungen auf. In Bezug auf bildbasierte Inhalte wird das immer dann in aller Deutlichkeit wahrnehmbar, wenn marginalisierte Subjekte aufgrund ihrer Körper(-teile) in ihrer Sichtbarkeit eingeschränkt werden, weil sie beispielsweise zu queer, zu nackt, zu trans*, zu sexy,[87] zu dick, zu be_hindert,[88] zu weiblich, zu haarig,[89] zu Schwarz[90] oder einfach zu unattraktiv (s. o.) scheinen, um anderen Nutzer*innen eine reibungslose Scroll-Erfahrung zu garantieren. Weniger offensichtlich, aber nicht weniger bemerkenswert sind jedoch zwei weitere Mechanismen, die mit den von Instagram (und mit großer Wahrscheinlichkeit auch von anderen Plattformen) eingesetzten *aesthetics ratings* zusammenhängen.

Zum einen werden (Un-)Sichtbarkeiten in Social-Media-Feeds, insbesondere auf der Bildebene, bei Weitem nicht mehr nur über Ausschlüsse und Zensur nicht-normativer Einzelfälle reguliert, sondern in viel größerem Maße dadurch, dass *normative* Inhalte durch Plattformmechanismen gefördert und begünstigt werden. In einer vergleichsweise wenig besprochenen Gegenbewegung zu viel diskutierten Verfahren der Zensur und des *shadowbanning* belohnen Plattformen tagtäglich normative Körper-Bilder, beispielsweise in Hinblick auf Geschlechterperformances.[91] Zum anderen greift, wie Instagrams *aesthetics ratings* erahnen lassen, die Regulierung viel tiefer, als es die plattformseitigen Zensurpraktiken gegenüber nicht erwünschten Körperbildern vermuten lassen. Die *regimes of recognition* betreffen längst nicht mehr nur die Sichtbarkeiten der konkreten Subjekte selbst – sie bestimmen viel umfassender, welche und wessen Bildwelten, Geschmäcker, Ästhetiken, Styles und Themen auf welche Arten gezeigt und gesehen werden können. Jacobsen hält fest: «Regimes of recognition on algorithmic media not only have the capacity to render certain people nonrecognisable; they also have the power to circumscribe people's parameters of attention and perception more generally.»[92]

Algorithmisch kuratierte Social-Media-Feeds beeinflussen nicht nur, *wer* gesehen wird, sondern, *wie* und *was* gesehen werden kann und somit Repräsentation erfährt. Damit legen Plattformen viel grundsätzlicher die Bedingungen dafür fest, wie wir als Subjekte für andere und uns selbst sichtbar oder unsichtbar gemacht und *erkannt* werden können[93] – was wiederum eine maßgebliche Voraussetzung für unsere soziale Existenz ist, wie Jacobsen betont: «Recognition is crucial for how we act and interact with others in society.»[94]

In Social-Media-Feeds hängt *recognition* zu großen Teilen davon ab, welche und wessen Bilder und Geschmäcker sichtbar sind und Repräsentation erfahren. Bestimmt wird das zu großen Teilen von den ökonomisch getriebenen Interessen der Plattformbetreiber*innen, die massiv Einfluss darauf nehmen,

87 Vgl. Brooke Erin Duffy, Colten Meisner: Platform Governance at the Margins. Social Media Creators' Experiences with Algorithmic (In) Visibility, in: *Media, Culture & Society*, Bd. 45, Nr. 2, März 2023, 285–304, doi.org/10.1177/01634437221111923; Carolina Are: The Assemblages of Flagging and De-Platforming against Marginalised Content Creators, in: *Convergence. The International Journal of Research into New Media Technologies*, Bd. 30, Nr. 2, Apr. 2024, 922–937, doi.org/10.1177/13548565231218629.

88 Vgl. Chris Köver, Markus Reuter: Diskriminierende Moderationsregeln: TikToks Obergrenze für Behinderungen, *netzpolitik.org*, 2.12.2019, netzpolitik.org/2019/tiktoks-obergrenze-fuer-behinderungen (4.12.2025).

89 Vgl. Arvida Byström, Molly Soda, Chris Kraus (Hg.): *Pics or It Didn't Happen. Images Banned from Instagram*, München u. a. 2016.

90 Vgl. Jacobsen: Regimes of Recognition on Algorithmic Media, 3646; Duffy, Meisner: Platform Governance at the Margins, 297.

91 Vgl. Ysabel Gerrard, Helen Thornham: Content Moderation. Social Media's Sexist Assemblages, in: *New Media & Society*, Bd. 22, Nr. 7, Juli 2020, 1266–1286, hier 1280, doi.org/10.1177/1461444820912540.

92 Jacobsen: Regimes of Recognition on Algorithmic Media, 3650.

93 Vgl. ebd., 3653.

94 Ebd., 3642.

welche Bilder als *verteilbar* gewertet und in ihrer Sichtbarkeit subventioniert werden – und welche nicht. Durch die Automatisierung dieser Eingriffe mithilfe von *aesthetics ratings* werden Bilddatenbanken wie LAION-Aesthetics, AVA und SAC zu maßgeblichen Akteur*innen in normativen Feedback-Loops zwischen den *sur-* und *subfaces* von Plattformumgebungen. Die in sie eingeschriebenen Geschmacksurteile und Repräsentationen beeinflussen auf mannigfaltige Weise, wie und welche Bilder im Plattformkapitalismus gesehen werden können.