

# DOES THIS PERSON EXIST?

KI-GENERIERTE PORTRÄTS UND IHRE PREKÄRE EXISTENZ  
IM DIGITALEN RAUM

Paul Werling

21: INQUIRIES INTO ART, HISTORY, AND THE VISUAL  
#4-2023, S. 745–781

<https://doi.org/10.11588/xxi.2023.4.101216>



ABSTRACT: DOES THIS PERSON EXIST? AI GENERATED  
PORTRAITS AND THEIR PRECARIOUS EXISTENCE IN THE  
DIGITAL

In this photo-theoretical article I interrogate portraits generated by NVIDIA's *StyleGAN2*. Even though these portraits are completely artificial and the pictured persons have never existed, they are surprisingly realistic and could even be read as photographs. I therefore ask the question: How do these pictures relate to the photographic and can we consider them as a new aspect of the medium? I will also present an attempt to locate these images on the Internet and discuss how they are used today. The result of this experiential localization shows that the pictures are mostly used in dubious and even criminal ways.

KEYWORDS

[www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com); Generierte Porträts; Digitale Bildlichkeit; Fototheorie; Künstliche Intelligenz; Maschinelles Lernen; StyleGAN2; Deep-Fake; Deep-Learning.

Digital generierte Porträts sind im Begriff, ein neues Kapitel in der Kulturgeschichte des Gesichts aufzuschlagen. Was vermutlich vor einigen Jahren noch als ferne Zukunftsvision abgetan worden wäre, ist durch gewaltige Fortschritte auf dem Forschungsfeld der künstlichen Intelligenz möglich geworden: maschinell erzeugte Bilder, die dem Fotografischen in einem Maße entsprechen, dass sie die gesellschaftlichen Versprechen und Handlungsweisen um das Medium auf die Probe stellen. Möglich ist das durch die 2019 von NVIDIA entwickelte künstliche Intelligenz *StyleGAN2*. Den öffentlichen Quellcode dieser KI nutzend veröffentlichte der Softwareentwickler Philip Wang noch im selben Jahr die Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Auf dieser Website präsentiert sich das verblüffende Potential der KI in einer endlosen Galerie. Sie ist in der Lage, ohne referentiellen Wirklichkeitsbezug Gesichter zu generieren, die in vielen Fällen täuschend echt aussehen [Abb. 1]. Medial an die Fotografie angelehnt, sollen die Bilder als solche gelesen werden. Aufgenommen wurde die Technologie mit Begeisterung und Besorgnis gleichermaßen. Die Tatsache, dass auf diese Weise entstandene Gesichter einen solchen Realitätsgrad erreichen können, sorgt für Verunsicherung. Zudem wird kritisiert, dass sowohl Bilder als auch Code frei verfügbar sind.<sup>1</sup>

Kritisch soll im Folgenden das Phänomen computergenerierter Porträts aus mehreren Perspektiven betrachtet werden. Zunächst gebe ich einen Einblick in die ersten bekannten Missbrauchsfälle und frage danach, wie diese Bilder detektiert werden können. In einem experimentellen Ansatz werde ich sodann mit einer Software zur Gesichtserkennung versuchen, solche Bilder zu lokalisieren, um nach ihrer spezifischen Verwendung zu fragen. Bereits jetzt sei vorweggenommen: Die Ergebnisse zeigen, dass alle erfolgreich lokalisierten Bilder in Zusammenhang mit Betrugsmaschinen stehen. Hieran knüpfen sich weitreichende Fragen, auf die Antworten gegeben werden sollen: Was sind die Grundprinzipien dieser Technologie? Wie lässt sich die rezeptionsästhetische Erfahrung theoretisieren? Und wie sind sie mit dem Fotografischen in Beziehung zu setzen?

## I. Die ersten Fälle missbräuchlicher Nutzung

Im März 2021 wandte sich das FBI mit einer Warnung an die Öffentlichkeit. In ihrem Bericht *Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations* sammelt die Behörde bekannt gewordene Fälle krimineller oder zumindest verdächtiger Identitätsverschleierung und spricht eine deutliche Warnung aus:

Malicious actors almost certainly will leverage synthetic content for cyber and foreign influence operations in the

<sup>1</sup>

Adam Ghahramani, Why ThisPersonDoesNotExist (and its copycats) need to be restricted, in: *VentureBeat*, 03.03.2019 (30.09.2023).



[Abb. 1]

Ein von *StyleGAN2* generiertes Porträt auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus (*StyleGAN2*), digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](http://www.thispersondoesnotexist.com), Creative Commons Licence (05.12.2023).

next 12–18 months. Foreign actors are currently using synthetic content in their influence campaigns, and the FBI anticipates it will be increasingly used by foreign and criminal cyber actors for spearphishing and social engineering in an evolution of cyber operational tradecraft.<sup>2</sup>

Diese Warnung bezieht sich dabei exklusiv auf den manipulativen Einsatz von „techniques based on artificial intelligence (AI) or machine learning technologies“<sup>3</sup>, also jenen Bildern, die im Folgenden genauer betrachtet werden sollen.

Als 2019 die Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) die Bilder erstmals der Öffentlichkeit zugänglich machte, wurden neben begeisterten auch kritische Stimmen laut. Von Anfang an wurde die Befürchtung geäußert, dass die Technologie im politischen Kampf auf digitalen Plattformen missbraucht werden könnte und zudem Betrüger:innen in die Hände spielen dürfte.<sup>4</sup> Noch im selben Jahr löschte Facebook ein Netzwerk von Fake-Profilen und Gruppen, das insgesamt mit 55 Millionen Nutzer:innen vernetzt war.<sup>5</sup> Dieses Netzwerk war multinational organisiert und produzierte Content, der von harmlosen *Click-Bait*-Artikeln bis zu politischer Propaganda eine breite Spannweite abdeckte. Die amerikanischen Seiten der Organisation fielen insbesondere durch Unterstützung des damaligen Präsidenten Donald Trump auf und wurden mit der umstrittenen amerikanischen Medienfirma „The Beauty of Life“<sup>6</sup> verknüpft. Die Besonderheit dieses Netzwerkes war die Nutzung von technisch generierten Profilbildern bei einer Vielzahl der Fake-Profile und bedeutet damit den ersten entdeckten Missbrauchsfall dieser Technologie.<sup>7</sup> Unter den von Meta entdeckten *Coordinated Inauthentic Behavior*-Profilen, also jenen, die koordiniert und zentral organisiert zu täuschen versuchen, machen die mit generierten Porträts verschleierte Profile 2019 und 2020 nicht einmal 5%

2

Der Bericht des FBI verweist auf einige Fallanalysen der Medienanalyse Firma Graphika. Diese dokumentieren die erwähnten Fälle detailliert. Federal Bureau of Investigation (FBI), Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations, in: *FBI*, 10.03.2021 (30.09.2023).

3

Ebd.

4

Ghahramani, Why ThisPersonDoesNotExist.

5

Nathaniel Gleicher, Removing Coordinated Inauthentic Behavior from Georgia, Vietnam and the US, in: *Meta*, 20.12.2019 (30.09.2023). Siehe auch: Ben Nimmo, C. Shawn Eib, L. Tamora, Kate Johnson, Ian Smith, Eto Buziashvili, Alyssa Kann, Kanishk Karan, Esteban Ponce de León Rosas and Max Rizzuto, #OperationFFS: Fake Face Swarm. Facebook Takes Down Network Tied to Epoch Media Group That Used Mass AI-Generated Profiles, in: *Graphika*, 20.12.2019 (30.09.2023), 2.

6

Jordan Liles, Exclusive. Expanding Pro-Trump Outlet ‘The BL’ Is Closely Linked to The Epoch Times, in: *Snores*, 11.10.2019 (30.09.2023).

7

Nimmo et al., #OperationFFS, 3.

der gesperrten Profile aus. 2022 jedoch, so der Jahresbericht von Meta, sind zwei Drittel der gesperrten CIB-Profilen mit generierten Gesichtern getarnt. Entscheidend ist hierbei, dass die Profile durch ihre Aktivität und nicht anhand der Bilder enttarnt werden.<sup>8</sup> Diese können nämlich auch vier Jahre nach ihrer Geburtsstunde nicht technisch bestimmt werden, sondern müssen weiterhin von trainierten Expert:innen identifiziert werden. Der geschulte Blick bleibt bislang der zuverlässigste Identifikator.<sup>9</sup>

Anhand einiger Bildbeispiele zeigen die Medienanalytist:innen von Graphika in ihrer Vorfalleanalyse auf, wie sich solche Bilder identifizieren lassen.<sup>10</sup> Denn obwohl die Porträts auf den ersten, uninformierten Blick sehr überzeugend wirken, lassen sich doch wenigstens bei einigen von ihnen auch anhand digitaler Artefakte Hinweise auf ihre Entstehung finden. Asymmetrische Brillengestelle sowie eigenartig gemorphte Schmuckelemente und Hintergründe, die mit der physischen Welt nichts gemein haben, sind dabei die besten Indikatoren [Abb. 2, Abb. 3, Abb. 4 und Abb. 5]. Dennoch äußern die Forscher:innen Bedenken über „the ease with which the operation managed to generate so many synthetic pictures in order to give its fake accounts (mostly) convincing faces.“<sup>11</sup> Denn auch wenn mit geschultem Blick ein solches Bild in vielen Fällen identifiziert werden kann, wird einem alltäglichen Sehen die eigentliche Natur dieser Bilder wohl meist entgehen. Wie oft kontrollieren User:innen die kleinen Profilbilder auf ihren technischen Ursprung? Ohne Kenntnis von dieser noch neuen Technologie ist eine solch vorsichtige Praxis wahrscheinlich die Ausnahme und die Bilder verschwinden in der Masse authentischer Profilbilder.

Doch welchen Vorteil bieten solche Bilder für manipulative Akteur:innen, wenn sie nach gegebenem Stand der Technik entlarvt werden können? Profilbilder in den Sozialen Medien haben vor allem einen Zweck: Sie sind Bindeglied zwischen digitaler Identität und realer Person. Profilen, die ihr Klargesicht zeigen, wird intuitiv mehr Authentizität und Autorität zugeschrieben als solchen, die sich in der Anonymität verstecken. Zumeist wird die Kontrolle der Echtheit mit einem schnellen Blick auf das Bild abgehakt. Kollektiv mussten wir vor 2019 nicht lernen, ein digitales Gesicht kategorisch in Frage zu stellen. In organisierten Propaganda- und Fake-News-Kampagnen im Kampf um politische Einflussnahme eröffnet sich auf diese Weise ein neues Potential für manipulative Akteur:innen. Die generierten Bilder bieten zwei entscheidende Vorteile gegen-

8

Ben Nimmo und David Agranovich, Recapping Our 2022 Coordinated Inauthentic Behavior Enforcements, in: *Meta*, 15.12.2022 (30.09.2023).

9

Shannon Bond, AI-generated fake faces have become a hallmark of online influence operations, in: *NPR*, 15.12.2022 (30.09.2023).

10

Nimmo et al., #OperationFFS, 21.

11

Ebd.

Does this Person Exist?



[Abb. 2]

Ein von *StyleGAN2* generiertes Porträt auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Die ungleich generierten Brillenbügel zeigen eine typische Schwäche des GANs. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus (*StyleGAN2*), digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](http://This Person Does Not Exist), Creative Commons Licence (05.12.2023).



[Abb. 3]

Ein von *StyleGAN2* generiertes Porträt auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Die digitalen Artefakte im Hintergrund zeigen eine typische Schwäche des GANs. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus (*StyleGAN2*), digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](http://This Person Does Not Exist), Creative Commons Licence (05.12.2023).

Does this Person Exist?



[Abb. 4]

Ein von *StyleGAN2* generiertes Porträt auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Das Gesichtsartefakt am linken Bildrand ist eine typische Schwäche des GANs. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus (*StyleGAN2*), digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](http://This Person Does Not Exist), Creative Commons Licence (05.12.2023).



[Abb. 5]

Ein von *StyleGAN2* generiertes Porträt auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com). Die schlecht gerenderten und zudem ungleichen Ohrringe sind ein typischer Fehler des *GANs*. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus (*StyleGAN2*), digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website *This Person Does Not Exist*, Creative Commons Licence (05.12.2023).

über gestohlenen Bildern von Realpersonen: Sie sind bei rückwärts gerichteten Bildersuchen keiner Quelle zuzuordnen und da sie keiner Realperson zugehörig sind, können die Profile nicht wegen Identitätsdiebstahls gemeldet werden. Sie sind digitale Masken, die sich einer Referenz verweigern,<sup>12</sup> zugleich aber vorgeben, genau diesem referentiellen Bildtypus – dem Fotografischen – zu entsprechen. Im Wissen um die politische Gefährlichkeit von Massenmanipulation auf den digitalen Plattformen ist es deshalb unbedingt notwendig, technische Detektionssysteme zu entwickeln. Diese oftmals täuschend echt generierten Bilder sind mit Sicherheit nicht das Ende der Entwicklung, sie verkörpern vielmehr ihr anfängliches Potential. Es wird der Zeitpunkt kommen, an dem selbst der informierte Blick die Unterscheidung zwischen generierten Bildern und solchen, die imitiert werden sollen, nicht mehr gewährleisten kann und die Bestimmung nur noch mit technischen Mitteln durchgeführt werden kann.<sup>13</sup>

## II. Wie lassen sich generierte Porträts finden?

In Frage steht also, wie sich generierte Porträts mit unseren heutigen Mitteln im digitalen Raum finden lassen. Fast alle bisher bekannt gewordenen Fälle wurden auf Twitter, Facebook und Instagram entdeckt. Die dort operierenden Fake-Netzwerke wurden jedoch nicht anhand der generierten Bilder enttarnt, sondern weil ihre verdächtigen Aktivitäten automatisiert im System detektiert wurden.<sup>14</sup> Hier wird deutlich, dass es bisher keine technologischen Lösungen gibt, um die Bilder als solche kategorisch zu identifizieren. Ein weiteres Problem bei der Recherche bereits bekannter Vorkommen ist die uneinheitliche Benennung der Bilder. Besonders der häufig genutzte Begriff *Deep Fake* ist unscharf, denn er bezeichnet eine Vielzahl von Technologien, die zumeist auf die Manipulation bestehender Medien spezialisiert sind.

Ein niederschwelliges Tool bei der Bildersuche ist die rückwärtsgewandte Bildersuche von Google. Diese Technik ist für gewöhnlich sehr treffsicher, führt aber bei den generierten Bildern nicht zu den gewünschten Ergebnissen. Statt ähnlich generierter Bilder schlägt Google stattdessen reale Personen vor, die den generierten zumindest ähnlich sind. Erfolgreich wäre diese Methode

<sup>12</sup>

Hans Belting, *Faces. Eine Geschichte des Gesichts*, München 2019, 296.

<sup>13</sup>

Es wird bereits daran geforscht, derartige Bilder technisch zu detektieren. Ein vielversprechender Versuch scheint die Frequenzanalyse zu sein, denn die GANs hinterlassen Produktionsspuren, die im *discrete cosine transformation (DCT)*-Spektrum sichtbar werden. Dies ist jedoch eine komplexe technische Anwendung, die meine Fähigkeiten übersteigt und zudem viele Bildbeispiele erfordert. Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa und Thorsten Holz, Leveraging Frequency Analysis for Deep Fake Image Recognition, in: *arXiv*, 2020.

<sup>14</sup>

Gleicher, Removing Coordinated Inauthentic Behavior. Siehe auch: Siddharth Venkataramkrishnan, After deepfakes, a new frontier of AI trickery. Fake faces, in: *Financial Times*, 12.10.2020 (30.09.2023).

jedoch nur, wenn das exakt gleiche Bild bereits im Internet aufgetaucht wäre. Da [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) keine öffentlichen Bildarchive zur Verfügung stellt, ist die Wahrscheinlichkeit, auf diese Weise ein Bild zu finden, verschwindend gering. Eine effektive Methode, die gesuchten Bilder zu finden, entdeckte ich eher durch Zufall. In der Einleitung seines Essays *Gesichtserkennung* beleuchtet der Bildwissenschaftler Roland Meyer zwei Unternehmen, die die bis dato effektivste Software zur Gesichtserkennung entwickelt haben: Clearview AI und PimEyes.<sup>15</sup> Einer derart fortgeschrittenen Identifikationstechnologie wohnt immenses Potential zum Missbrauch und zur Massenüberwachung inne, zurecht stehen die Unternehmen in der Kritik. Mit ihnen zeichnet sich das Ende der ohnehin schon marode gewordenen Anonymität in physischen und digitalen Räumen ab.<sup>16</sup> Hinzu kommt, dass Clearview AI und PimEyes die Trainingsdaten für ihre *Deep-Learning* Prozesse illegal von den gängigen Social Media Plattformen gestohlen haben.

Insbesondere das Unternehmen PimEyes steht massiv in der Kritik. Denn während Clearviews Software nicht einfach im Internet zugänglich ist (die Firma wendet sich an größere Akteure), hat PimEyes ein Suchportal erstellt, in dem jede:r ohne Zugangsbeschränkung Gesichter suchen kann. Die Suchmaschine durchforstet in Sekundenbruchteilen das Internet und listet zuverlässig andere Websites auf, auf denen das gesuchte Gesicht auftaucht.<sup>17</sup> Eine Einschränkung jedoch besteht: Ohne zu bezahlen, sieht man nur Thumbnails der Ergebnisse, erhält aber keinen Zugang auf die jeweilig verlinkten Websites. Trotz erheblicher Vorbehalte gegenüber PimEyes entschied ich mich, in einem Versuch generierte Porträts in die Suchmaske hochzuladen.<sup>18</sup>

In Erwartung ähnlich schlechter Ergebnisse wie bei der Google-Suche galt mein hauptsächliches Interesse der Frage, ob die Anwendung generierte Bilder erkennen kann und als solche markiert. Das wenig überraschende Ergebnis war, dass die Bilder nicht als künstlich enttarnt wurden. Überraschend waren jedoch die Suchergebnisse [Abb. 6 und Abb. 7]. Ganz eindeutig sind hier verschiedene Personen zu sehen. Mit Blick auf die sonstige Präzision der Suchmaschine ist das zumindest ungewöhnlich. Weiter

15

Roland Meyer, *Gesichtserkennung*, Berlin 2021, 6.

16

Daniel Laufer und Sebastian Meineck, PimEyes. Eine polnische Firma schafft gerade unsere Anonymität ab, in: [Netzpolitik.org](http://Netzpolitik.org), 10.07.2020 (30.09.2023).

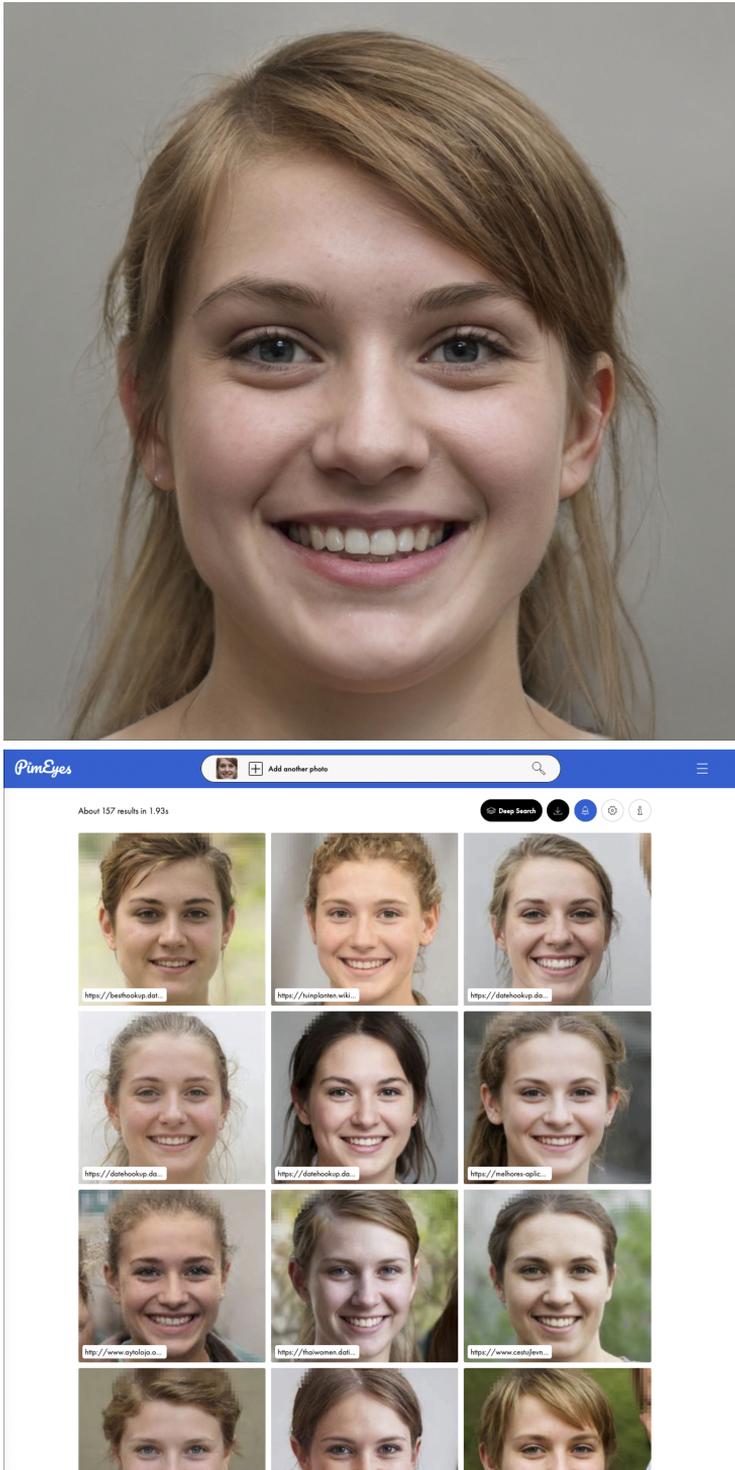
17

[Netzpolitik.org](http://Netzpolitik.org) hat dessen Zuverlässigkeit und hohe Trefferquote dokumentiert. In diesem Artikel geben die Autor:innen einen detaillierten Überblick über das Unternehmen. Es sei an dieser Stelle auch darauf hingewiesen, dass das ehemals polnische Unternehmen PimEyes aufgrund der Kritik und drohender rechtlicher Konsequenzen mittlerweile seinen Geschäftssitz auf die Seychellen verlagert hat. Siehe: Laufer und Meineck, PimEyes.

18

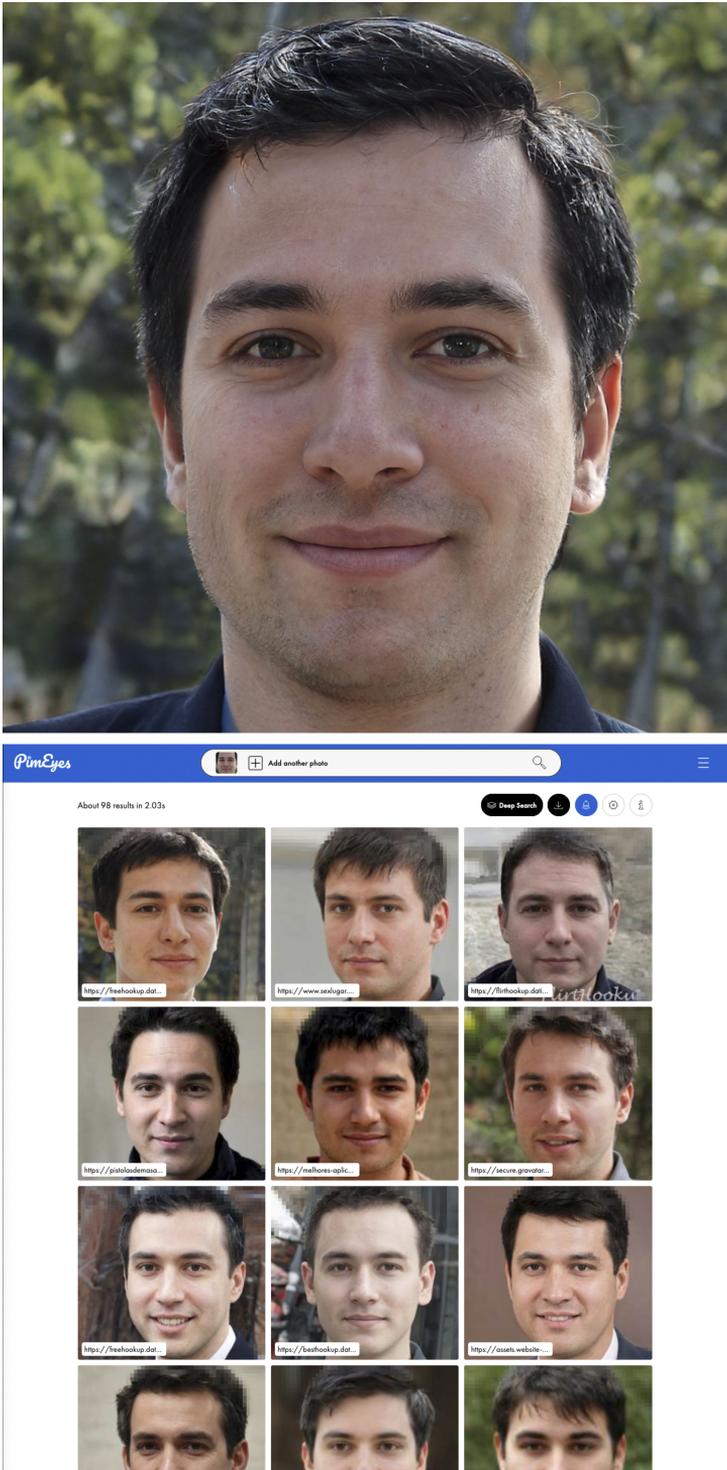
Alle Abbildungen sind der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) entnommen. Ausgewählt habe ich nur solche, die möglichst frei von Artefakten und anderen Indikatoren sind, die auf ihre Künstlichkeit hinweisen.

## Does this Person Exist?



[Abb. 6]

Screenshot von der Bildersuche auf der Website [PimEyes](#). Oben ist das gesuchte Porträt, unten ein Screenshot der Suchergebnisse. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus, digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](#), Creative Commons Licence (15.01.2022).



[Abb. 7]

Screenshot von der Bildersuche auf der Website [PimEyes](#). Oben ist das gesuchte Porträt, unten ein Screenshot der Suchergebnisse. Phil Wang, computergeneriertes Porträt, 2019, generativer Algorithmus, digitales Bild (JPEG), 1024 × 1024 Pixel, mit freundlicher Genehmigung der Website [This Person Does Not Exist](#), Creative Commons Licence (15.01.2022).

verbindet die verschiedenen Gesichter eine schwer mit Worten zu fassende Ähnlichkeit. Obgleich sie eindeutig individuelle Charakteristika haben und sich in allen Belangen unterscheiden, verbindet sie dennoch etwas Durchschnittliches – man gewinnt ein Gefühl von verwandter Konstruiertheit. Hinzu kommen ein streng quadratisches Bildformat und die Position der Augen an der immer gleichen Stelle im Bild. Die naheliegende Vermutung ist, dass es sich stets um generierte Gesichter handelt und die Software deshalb keine Unterschiede feststellen kann.

Um diese These zu überprüfen, entschied ich mich, ein bezahlpflichtiges Abo bei PimEyes abzuschließen.<sup>19</sup> Mit diesem Zugang war es möglich, die Bilder einzeln zu öffnen und manuell zu überprüfen. Dabei zeigte sich, dass alle auf den Screenshots [Abb. 6 und Abb. 7] sichtbaren Bilder die typischen Merkmale der generierten Bilder aufweisen. Neben den Artefakten (Ohringe, Brillen, Hintergrund-Artefakte) ist ein zuverlässiger Indikator die Position der Augen. [Abb. 8 und Abb. 9] zeigen die gefundenen Bilder transparent gestapelt. Hierbei wird deutlich, dass, obwohl die Gesichter verschieden ausgerichtet sind, die Augen (und zum großen Teil auch die Münder) auf der exakt gleichen Bildposition liegen.<sup>20</sup> Auch die Google-Suche nach den einzelnen Bildern brachte kein Ergebnis, das auf einen fotografischen Ursprung der Bilder hindeutet. Die Annahme, dass sich mit der PimEyes-Suchfunktion die von *StyleGAN2* generierten Bilder finden lassen, sehe ich hierin bestätigt.<sup>21</sup>

Daran anschließend stellt sich die Frage, mit welchem Zweck diese Bilder eingesetzt werden. Hierzu habe ich die Websites der achtzehn Gesichter auf den Screenshots [Abb. 6 und Abb. 7] aufgesucht. Zehn der Seiten waren Dating-Portale, vier Gesichter tauchten als Schreiber:innen von Produktreviews auf. Zwei der Porträts waren Profilbilder von Autor:innen, und zwei Bilder befanden sich auf Seiten, deren Zweck sich mir nicht erschloss.<sup>22</sup> Die Anonymisierung von Autor:innen unter Zuhilfenahme falscher und nun auch

19

Ich bin mir der Problematik dieser Entscheidung bewusst. Dadurch unterstütze ich ökonomisch ein Unternehmen, das im besten Falle fragwürdig, im schlimmsten Fall jedoch kriminell ist und eine immense Gefahr für Datensicherheit und Anonymität darstellt. Jedoch scheint dieses Unternehmen unbeabsichtigt die bisher einzige Möglichkeit entwickelt zu haben, generierte Bilder im Internet zu tracken und die damit verbundenen Praktiken zu dokumentieren. Deshalb habe ich mich entschieden, den Dienst für einen Monat zu abonnieren und anschließend wieder zu kündigen. Zudem waren alle von mir gesuchten Gesichter künstlicher Natur, somit wurden keine Persönlichkeitsrechte verletzt und das System nicht weiter an realen Gesichtern trainiert.

20

Da die Bilder verschieden gut aufgelöst waren, habe ich sie zunächst auf ihre native Auflösung 1024 × 1024 Pixel vergrößert.

21

Strenggenommen könnte es sich bei den Bildern auch um die präparierten Trainingsbilder von Flickr handeln, jedoch deuten die Artefakte daraufhin, dass mindestens der Großteil der Bilder generiert ist. Hierfür spricht auch, dass die Rückwärtssuche bei Google und Pim-Eyes nicht auf Flickr oder den Bild-Datensatz verweisen. Zudem bieten die generierten Bilder mehr Vorteile im Anbetracht des manipulativen Bildgebrauchs, den ich folgend aufzeige.

22

Die Links zu den Bildern sind im Anhang in Reihenfolge der Abbildungen aufgelistet.



[Abb. 8]  
Die digital gestapelte Ansicht der ersten neun Suchergebnisse [Abb. 6]. Die übereinstimmende Positionierung von Augen und Mündern ist ein starkes Indiz auf den Ursprung der Bilder.

Does this Person Exist?



[Abb. 9]  
Die gestapelte Ansicht der ersten neun Suchergebnisse [Abb. 7]. Die übereinstimmende Positionierung von Augen und Mündern ist ein starkes Indiz auf den Ursprung der Bilder.

generierter Bilder ist eine Entwicklung, die bereits eingangs zur Sprache kam. Wie die Recherchen von Graphika gezeigt haben, gibt es globale Netzwerke, die sich auf anonymisierte Content-Produktion spezialisiert haben. Wie schon beschrieben, haben diese Inhalte eine Spannweite von harmlos bis zu Fake News und Propaganda.<sup>23</sup> Fake Reviews sind ebenfalls ein bekanntes Problem.<sup>24</sup> Es überrascht nicht, dass auch in diesem Bereich auf generierte Bilder zurückgegriffen wird, denn auch hier generiert ein Profil mit verknüpftem Gesicht mehr Vertrauen als ein anonymes.

Das interessanteste Ergebnis dieses Experiments betrifft jedoch die Dating-Portale. Seit Jahren haben große Anbieter wie Tinder mit Fake-Profilen zu kämpfen.<sup>25</sup> Üblicherweise werden diese Profile durch Bots bespielt, deren Ziel es ist, die User:innen auf andere Plattformen zu locken oder ihnen Daten zu entlocken. Die Identifikation solcher Profile ist nicht immer einfach, denn sie nutzen in der Regel unrechtmäßig kopierte Bilder realer Personen. Abseits der großen Plattformen, die streng gegen solche Täuschungsversuche vorgehen, gibt es kleinere Dating-Plattformen, die wiederum Betrug zu ihrer Masche gemacht haben. Im März 2021 veröffentlichte die Verbraucherzentrale einen Bericht, der Betrug im Online-Dating dokumentierte.<sup>26</sup> Darin wurden 187 Plattformen identifiziert, die Fake-Profile einsetzen. Diese Fake-Profile werden von Mitarbeiter:innen oder Cloud-Diensten betrieben. Ziel ist es, potentielle Kund:innen, also Menschen, die auf der Suche nach eine:r Partner:in sind, so lange wie möglich in ein Gespräch zu verwickeln, um sie so in bezahlpflichtige Abos und Nachrichten zu drängen. Die Kund:innen wissen dabei nicht, dass sie mit professionellen Chat-Schreiber:innen kommunizieren und betrogen werden.<sup>27</sup> Es liegt der Verdacht nahe, dass auf den bei PimEyes aufgetauchten Seiten diese oder ähnliche Betrugsmaschen vollzogen werden. Zwar tauchen die Bilder auf verschiedenen Seiten auf, aber die Namen sind auffallend ähnlich: *freehookup.dating*, *flirTHOOKUP.dating*, *besthookup.dating*, *datehookup.dating* und *thaiwoman.dating*. Offenkundig handelt es sich hier um ein Netzwerk.

23

Nimmo et al., #OperationFFS.

24

Nicole Nguyen, Fake Reviews and Inflated Ratings Are Still a Problem for Amazon, in: *The Wall Street Journal*, 13.06.2021 (30.09.2023).

25

Mila Krull, Gefahr durch Bots bei Dating-Apps. Wie man Fake-Accounts erkennt, in: *RedaktionsNetzwerk Deutschland*, 24.07.2020 (30.09.2023).

26

Verbraucherzentrale, Marktüberblick Online-Dating-Portale. Einschätzung zu einem problembehafteten Markt, in: *Verbraucherzentrale*, 2021 (30.09.2023).

27

STRG\_F hat eine Doku gedreht, in der die Journalistin undercover einen solchen Job annimmt und ihre Bedingungen und Aufgaben dokumentiert: Nadia Kailouli und David Diwiak, Undercover als Chatschreiberin. Falsche Flirts auf Dating-Plattformen. STRG\_F, *Youtube*, 02.06.2020 (13.12.2023). Siehe auch: Leonard Scharfenberg, Betrug auf Datingseiten. Das Geschäft mit gebrochenen Herzen, in: *TAZ*, 30.10.2020 (30.09.2023).

Alle diese Websites sind nachlässig programmiert und – wenigstens für ein aufmerksames Auge – voll augenscheinlicher Fake-Profile. Versucht man sich anzumelden, wird man auf andere, zum Teil pornografische Websites weitergeleitet. Spätestens hier dürfte klar sein, dass diese Websites eine Betrugsmasche verfolgen und mit generierten Porträts überzeugende Profile zu erzeugen versuchen. Die fehlerhaften Userinterfaces und die stetige Weiterleitung auf ähnliche Websites deuten jedoch darauf hin, dass es primär um Datentracking und potenziell um Vireninfection geht. Es kann hier also ein halbwegs positives Zwischenfazit gezogen werden: Mit Gesichtserkennungssoftware lassen sich die generierten Porträts im Internet aufspüren. Weiter hat sich gezeigt, dass diese in betrügerischen Maschen eingesetzt werden, jedoch zum gegebenen Zeitpunkt nicht sehr überzeugend sind. Allerdings mahnt die Recherche der Verbraucherzentrale zur Vorsicht. Es dürfte nur eine Frage der Zeit sein, bis auch im professionelleren Dating-Betrug generierte Bilder auftauchen.

### III. Zur technischen Natur generierter Bilder

Um die Natur generierter Bilder zu verstehen, ist es unumgänglich, ihren Produktionsprozess zu befragen. Im Folgenden sollen kurz wesentliche Prinzipien maschinellen Lernens umrissen werden. Grundsätzlich kann zwischen zwei Lernprozessen unterschieden werden. Zum einen gibt es das *Machine Learning*. Hier sind die Trainingsdaten strukturiert und gelabelt, wodurch das System die gesuchte Struktur erkennen kann. Aufbauend auf das Feedback des menschlichen Operators ist das System in der Lage, strukturelle Optimierungen durchzuführen. Ziel ist es dabei, ein System zu bilden, das die erlernten Strukturen auf unbekannte Datensätze übertragen kann.<sup>28</sup> Zum anderen wurde das *Deep Learning* als Untersystem des *Machine Learnings* entwickelt. Im Bereich der Bild- und Spracherkennung hat dieses System in den letzten Jahren enorme Entwicklungen erfahren, und auch die hier untersuchten künstlich generierten Porträts beruhen auf diesem Prinzip.

*Deep Learning* baut auf einem *Artificial Neural Network (ANN)* auf. Grundsätzlich sind *Artificial Neural Networks* aus „simplen Recheneinheiten, den *Neuronen* sowie gerichteten, gewichteten Verbindungen zwischen diesen“<sup>29</sup> aufgebaut. Diese Neuronen sind in aufeinander folgenden und hierarchisch vernetzten *Layers* angeordnet. Die *Input-* und *Output-Layers* sind dabei einsehbar, die sogenannten *Hidden-Layers* dazwischen sind die Blackbox des Prozesses. Die nicht einsehbare Eigengestaltung des Arbeitsprozesses ist

<sup>28</sup>

Lorena Jaume-Palasi und Fritz Pieper, KI und Algorithmen, in: *Telemedicus*, 23.01.2019 (30.09.2023).

<sup>29</sup>

David Kriesel, *Ein kleiner Überblick über Neuronale Netze*, 2005, 35.

das, was *Deep Learning* auszeichnet.<sup>30</sup> Ist es nun die Aufgabe der KI, Bilder zu klassifizieren – beispielsweise zu entscheiden, ob die Fotografie einen Hund zeigt oder nicht – muss eine riesige Menge an Trainingsdaten zusammengetragen und ins System implementiert werden.<sup>31</sup> Dabei wird jedem Pixel des eingespeisten Bildes ein Neuron des *Input-Layers* zugeordnet. Eigenständig werden die Pixelinformationen (RGB-Werte und Belichtung) an die folgenden Neuronen-Schichten weitergegeben mit dem Ziel, Strukturen und Muster zu erkennen, die die Form eines Hundes verlässlich klassifizieren können.<sup>32</sup> Durch Feedback des:der Operator:in lernt das System, welche Strukturen den Hund klassifizieren und gewichtet die neuronalen Verbindungen entsprechend der korrekten Klassifizierung. Sobald die Trainingsdaten eine ausreichende Zuverlässigkeit gewährleisten, ist das System in der Lage auch unbekannte Bilder zu klassifizieren.<sup>33</sup> Dieses System wird als Diskriminator bezeichnet.

Dem gegenüber steht das *Deep Generative Learning*, der Generator. In diesem Anwendungsbereich versucht die KI nicht die Daten zu klassifizieren, sondern ihre Verteilung zu analysieren: „It tries to understand the distribution of data points, providing a model of how the data is actually generated in terms of a probabilistic model.“<sup>34</sup> Die Trainingsdaten sind dabei ungelabelt, so dass die statistischen Verteilungen der Datenpunkte aus dem Datenset extrahiert werden müssen. Dadurch ist das System in der Lage, Aussagen über die eingespeisten Daten hinaus zu treffen.<sup>35</sup> Bestehen die Trainingsdaten beispielsweise aus 100.000 Bildern von Hunden, lernt die KI, welche Datenpunkte beziehungsweise Pixel welche Information beinhalten müssen, um die eingespeiste Struktur (in meinem Beispiel den Hund) üblicherweise zu bilden. Darauf aufbauend ist das *ANN* bei erfolgreichem Training in der Lage, eine auf statistischer Verteilung beruhende Aussage zu extrapolieren. In diesem Fall wäre das ein von Grund auf generiertes Bild eines Hundes.

Auf sich allein gestellt ist das *Deep Generative Learning* jedoch fehleranfällig. Um Fehlerquoten entgegenzuwirken, veröffentliche

30

Michael A. Nielsen, *Deep Learning*, in: *Determination Press*, 2015 (27.11.2023).

31

Sven Behnke und Hannes Schulz, *Deep Learning. Layer-Wise Learning of Feature Hierarchies*, in: *Künstliche Intelligenz* 26, 2012, 357–363, hier 358.

32

Ebd.

33

Fritz Pieper und Oliver Stiemerling, Einführung in das Thema „Künstliche Intelligenz“, in: *Telemedicus*, 08.01.2019 (30.09.2023).

34

Satya Mallick, *Generative and Discriminative Models*, in: *LearnOpenCV*, 10.05.2021 (30.09.2023).

35

Mallick, *Generative and Discriminative Models*.

ten der amerikanische Informatiker Ian J. Goodfellow und sein Team 2004 ein System, das eine gesteigerte Zuverlässigkeit versprach. Das von ihm entwickelte *Generative Adversial Network (GAN)* stellt dafür ein generatives und ein diskriminierendes System in direkte Konkurrenz und erzeugt so ein qualitatives Lernsystem, das sich selbst stetig optimiert.<sup>36</sup> Beide *ANN* greifen auf den gleichen Datensatz zu. Aufgabe des Generators ist es, ausgehend von einem *Noise Input* eine Aussage zu erzeugen, die dem Datensatz möglichst ähnlich ist. Aufgabe des Diskriminators ist es zu entscheiden, ob die generierte Aussage dem Datensatz entspricht oder nicht. Gelingt es dem Generator nicht den Diskriminator zu überzeugen, implementiert der Generator das Feedback in sein *Neural Network*.

Eine Analogie, die Goodfellow vorschlägt, ist die des Kunstfälschers. Seine Intention ist es, ein Bild zu schaffen, das beispielsweise dem Stil nach den Bildern Vincent van Goghs zum Verwechseln ähnlich sieht, aber keine Kopie eines originalen Werkes ist. Der Kunstmarkt auf der anderen Seite kennt das *Œuvre* dieses Malers und muss kontrollieren, dass nur Originale im Umlauf sind. Der Fälscher ist erfolgreich, wenn er den Markt täuscht, der Markt ist erfolgreich, wenn er den Fälscher entlarvt. Durch Überlistung und Entlarvung findet ein beidseitiger Optimierungsprozess statt.<sup>37</sup>

Seit der ersten Veröffentlichung hat sich die Qualität der *GANs* immens verbessert. Während die ersten Outputs noch von geringer Auflösung waren, sind heutige Anwendungen im Stande Bilder bis zu der Auflösung  $1024 \times 1024$  Pixel zu generieren. Grund dafür sind zum einen die wachsende Masse an Trainingsdaten im Internet, zum anderen die Sprünge in der Rechenleistung von Computern. Mittlerweile gibt es eine Vielzahl verschiedener *GANs*.<sup>38</sup> Gegenstand der Betrachtung wird jedoch ausschließlich das von NVIDIA entwickelte und 2018 veröffentlichte *StyleGAN* sein.<sup>39</sup> Grundsätzlich ist dieses *GAN* in der Lage, jeden Bildtypus zu erzeugen, sofern ausreichende Bilder zum Trainieren vorliegen. Trainiert wurde das *GAN* beispielsweise an Bildern von Katzen, Pferden, Autos und Küchen. Jedoch sind diese Bilder weitaus weniger überzeugend als die von Gesichtern. Die Ursache dieser besonderen Qualität lässt sich in einer gewissen formalen Erwartbarkeit des menschlichen Gesichts vermuten. Während andere Bildsujets eine viel größere Bandbreite an Motiven miteinbeziehen, lässt sich das Gesicht auf

36

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville und Yoshua Bengio, Generative Adversarial Networks, in: *Advances in Neural Information Processing System* 3/11, 2014, 1–2.

37

Ebd., 1.

38

Ein kurzer Überblick findet sich hier: Jason Brownlee, 18 Impressive Applications of Generative Adversarial Networks (GANs), in: *Machine Learning Mastery*, 12.07.2019 (30.09.2023).

39

Timo Aila, Tero Karras und Samuli Laine, A Style-Based Generator Architecture for Generative Adversarial Networks, in: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43/12, 4217–4228, 2021.

sehr einfache Grundprinzipien reduzieren. Gesteigert wird dies in der Vorformatierung der 70.000 Gesichter umfassenden Trainingsdatenbank. Die von Flickr downgeloadeten Bilder werden automatisiert in das quadratische Bildformat beschnitten und die Positionierung der Gesichter angeglichen.<sup>40</sup> Auch wenn die Lizenzbedingungen von Flickr den Download der Bilder erlauben, wirft der Umgang ethische Fragen auf:

Machine learning systems are trained on images like these every day – images that were taken from the internet [...] without context or consent. They are anything but neutral. [...] A computer vision system can detect a face or a building but not [...] any of the social and historical context surrounding that moment.<sup>41</sup>

Solche Dekontextualisierung negiert das Individuelle und reduziert es auf einen Datensatz.

Any individual image could easily be substituted for another and the system would work the same. According to this worldview, there is always more data to capture from the constantly growing and globally distributed treasure chest of the internet and social media platforms.<sup>42</sup>

Auch für die in dieser Arbeit gezeigten Porträts ist diese Art der Dehumanisierung grundlegend. Die individuellen Porträts der Personen sind nun nicht viel mehr als ein Möglichkeitsraum, der die potenzielle Bandbreite der Bildgenerierungen absteckt. Jedoch ist es dieser Möglichkeitsraum, der die Technologie von Durchschnittsbildern unterscheidet. Die Idee, aus verschiedenen Porträts eine Durchschnittserscheinung zu gewinnen, stellte 1878 erstmals der britische Naturforscher und Schriftsteller Francis Galton vor. Der physiognomischen Idee aus der menschlichen Erscheinung Rückschlüsse auf Eigenschaften ableiten zu können folgend, entwickelte er die Kompositfotografie. Indem er mehrere Porträts in einer Mehrfachbelichtung zu einem verschmolz, versuchte er einen qualitativen Umgang mit dem fotografischen Medium zu finden. „Im Ergebnis, so Galtons These, sehe man den Durchschnitt der Gesichter: eine bildgewordene Statistik.“<sup>43</sup> Voraussetzung dafür war eine strikte Vorformatierung der Bilder, wobei hierin die

<sup>40</sup>  
[Trainingsdatenbank Flickr-Faces-HQ Dataset \(FFHQ\) \(27.11.2023\).](#)

<sup>41</sup>  
Kate Crawford, *Atlas of AI*, New Haven, CT 2021, 94.

<sup>42</sup>  
Ebd.

<sup>43</sup>  
Roland Meyer, *Operative Porträts. Eine Bildgeschichte der Identifizierbarkeit von Lavater bis Facebook*, Konstanz 2019, 106.

Gemeinsamkeit mit den GAN-Verfahren besteht. Hingegen unterscheiden sich die beiden Verfahren im finalen Bild: Während Galton eine Synthese aus bestehenden Bildern erzeugte, generiert das GAN ein von Grund auf neues Bild. Keinesfalls ist dieses als Durchschnitt aller Trainingsdaten zu sehen, sondern als eine Neuformulierung der statistischen Auswertung von Trainingsdaten. Die Trainingsdaten bilden die Bandbreite der Möglichkeiten, bestimmen jedoch nicht als Ganzes die Erscheinung des Outputs. Und dies ist, zumindest auf dem fotografischen Feld, die entscheidende Neuerung: Ausgehend von Francis Galton wurde in der Geschichte der Fotografie immer wieder versucht, Mischporträts zu erzeugen, diese konnten rein technisch jedoch nie mehr als die konkrete Summe der Ausgangsbilder sein, das rein fiktive Porträt ohne faktisches Ausgangsmaterial musste bisher ein händischer Prozess bleiben. Nun aber ist das fiktive Porträt als technisches Erzeugnis nicht mehr die Summe aller Teile, sondern ein willkürliches Produkt aus dem Möglichkeitsraum aller Teile.

#### IV. Diskriminator:in wider Willen

Um die von *StyleGAN2* generierten Bilder zu erleben, sei empfohlen, die Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) zu besuchen. In dieser Galerie prekärer Existenz fließt ein nicht endender Strom generierter Gesichter.<sup>44</sup> Das erste Antlitz, das mir dort begegnete, war das einer jungen Frau [Abb. 1].<sup>45</sup> Mein anfänglicher Eindruck war von Erschrecken geprägt. Die real anmutende Erscheinung konnte ich nur schwer mit meinem Bewusstsein von der Entstehungsweise des Bildes in Einklang bringen. In meiner so geweckten Faszination begann ich über sie zu mutmaßen: Ich las sie als junge Frau in meinem Alter, erfreute mich an ihrer Natürlichkeit und mutmaßte über ihre Lebensgeschichte – nur um all das schon im nächsten Augenblick zu dementieren. Sie ist keine Frau, nicht jung und hat auch kein freundliches Lächeln. Sie ist ein statistischer Output beruhend auf einer großen Bilddatenbank. Dennoch führt sie nun, nach zwei Jahren des Nachdenkens über sie, in meinen Gedanken eine Schattenexistenz.

Das Wissen um eine maschinelle Annäherung an die Bildlichkeit einer biologischen Komplexität und Individualität ohne die konkrete Einflussnahme menschlicher Schaffenskraft erzeugte eine Erfahrung, die im Analogen wohl als „punctum“<sup>46</sup> beschrieben wor-

<sup>44</sup>

Ein einfacher Test belegt, dass es sich tatsächlich um eine Galerie handelt und nicht bei jedem Aktualisieren der Seite ein neues Bild generiert wird. Aktualisiert man die Website exakt zeitgleich mit zwei Browsern, so werden beide Browser das gleiche Bild präsentieren. Es findet somit keine Bedienung einer KI durch die User:innen statt.

<sup>45</sup>

In der folgenden Ausführung beziehe ich mich primär auf dieses Bild.

<sup>46</sup>

Roland Barthes, *Die helle Kammer. Bemerkungen zur Photographie*, Frankfurt a. M. 2016 [1980], 36.

den wäre, im Digitalen bisher einer solchen Begrifflichkeit jedoch entbehrt. Erwähnenswert ist es deshalb, weil auch hier die Medialität Ausgangspunkt einer starken Rezeptionserfahrung ist, allerdings in gegensätzlicher Argumentation. Denn während Roland Barthes dieses Erleben stark an die vergangenheitsweisende Zeitlichkeit des analogen fotografischen Mediums bindet,<sup>47</sup> tritt hier der umgekehrte Fall ein: Die Materialität der generierten Bilder verweigert sich referentieller Zeitlichkeit, obgleich sie in vertrauter Form erscheint. Ich blicke in ein „virtuelles Gesicht, das sich jeder Referenz auf einen natürlichen Träger [...] entzieht.“<sup>48</sup> Diese notwendige Verweigerung des Realen erzeugte eine neue, zunächst unbehagliche Erfahrung, die sich nicht mit meinem bisherigen fotografischen Erfahrungshorizont vereinbaren lässt. Ein derartiges Unbehagen ist im Bereich der Robotik und grafischen Bildsimulation als *Uncanny-Valley* bekannt.<sup>49</sup> Der Theorie zufolge fällt die Akzeptanz simulierter Menschendarstellungen ab einem bestimmten Grad der Anthropomorphisierung rapide ab – das *Uncanny-Valley*. „An dieser Stelle kippt der Eindruck von Realismus plötzlich in sein Gegenteil um und es entsteht der Eindruck des Unheimlichen, des lebenden Toten.“<sup>50</sup> Wie beschrieben haben viele der vorliegenden generierten Bilder das Potential, mich als Rezipienten von der Klippe in das *Uncanny-Valley* zu stoßen. Dies ist jedoch, wie auch deutlich wurde, stark vom Kontext abhängig. Die Besonderheit der Bilder ist ihr hoher Grad an Realismus im fotografischen Sinne. Dieses Kippmoment in Unbehagen kann nur funktionieren, wenn das Bild als ein generiertes identifiziert werden kann. Begegnet ein solches Porträt uns beispielsweise auf Social Media oder auf einem Führerschein, ist die Identifikation seiner computergenerierten Natur schwierig. Die Existenz der und das Wissen um diese Technologie machen es notwendig, fortan immer die Rolle des Diskriminators einzunehmen. Hat der technische Diskriminator ein Bild als der Referenz entsprechend akzeptiert, macht er uns zu Kompliz:innen in dem ewigen Ringen um Entlarvung und Überlistung. Verblüffend ist, dass dieser wissend geschärfte Blick, zumindest bis Dezember 2022, als die zuverlässigste diskriminatorische Technik angesehen wurde, auch bei dem Tech-Giganten Meta.<sup>51</sup>

47

Barthes, Die helle Kammer, 90–91.

48

Belting, Faces, 296.

49

Masahiro Mori, Karl F. MacDorman [Übers.] und Norri Kageki [Übers.], The Uncanny Valley, in: *IEEE Robotics & Automation Magazine* 19/2, 2012, 98–100.

50

Markus Rautzenberg, *Bild und Spiel. Medien der Unwissenheit*, Paderborn 2020, 32.

51

Bond, AI-generated fake faces.

## V. Zur fotografischen DNA generierter Bilder

Was unter digitalen Bedingungen als fotografisches Bild zirkuliert, ist [...] das Ergebnis eines Messvorgangs, bei dem Halbleitersensoren Lichtimpulse in elektrische Signale umwandeln, die sich als diskrete Zahlenwerte speichern, übertragen und verrechnen lassen.<sup>52</sup>

Durch diesen technischen Digitalisierungsprozess entsteht ein in zweifacher Hinsicht ungreifbares Produkt: Zum einen ist das Bild in digitaler Codeform physisch ortlos und lässt sich beliebig oft verschieben und kopieren, ohne dabei physisch bewegt oder „sichtbar“ zu werden. Zum anderen ist seine Sichtbarwerdung zeitkritisch und – im Digitalen – nie beständig. Abhängig von der Bildwiederholfrequenz des Monitors wird es in entsprechenden Zeitabständen ständig neu erzeugt.<sup>53</sup> Das digitale Bild als solches ist also für den Menschen unsichtbar, denn ‚digital‘ meint in erster Linie die Art der Speicherung.

The ‚image‘ on the other hand, whether generated once as a print or generated 60–100 times a second to be projected though a monitor, is *never* digital and is *always* analogue, otherwise we humans could never see it because we cannot see voltage differences.<sup>54</sup>

In seinem Kern ist ein digitales Bild, sofern sichtbar gemacht, ein streng geometrisches Mosaik aus einzelnen Pixeln. Dieses Mosaik wird in einem langen Zahlencode gespeichert. Es ist dabei eine reine „Designentscheidung“<sup>55</sup>, dass die Bilder hierbei der visuellen Norm analoger Fotografie entsprechen, der Code selbst gibt das nicht zwingend vor und könnte auf unterschiedlichste Arten ausgelesen werden. Die erste Zahl des Codes gibt dem auslesenden Programm die Information über den normierten Datentyp, also wie die Zahlen zu interpretieren sind; die folgenden Zahlen sind die Informationen jedes Pixels in der Reihenfolge des jeweiligen Bildpunktes. In den meisten Fällen sind die Pixel-Informationen Zahlenwerte, die sich auf den RGB-Farbraum beziehen.<sup>56</sup> Dies gilt prinzipiell für

<sup>52</sup>

Meyer, Operative Porträts, 419–420.

<sup>53</sup>

Peter Berz, Bitmapped Graphics, in: Axel Volmar (Hg.), *Zeitkritische Medien*, Bamberg 2009, 127–154, hier 144.

<sup>54</sup>

Yanai Toister, Photography. Love’s Labour’s Lost, in: *Photographies* 12/1, 2019, 117–133, hier 127.

<sup>55</sup>

Daniel Rubinstein, Fotografie nach der Philosophie. Repräsentationsdämmerung, Leipzig 2020, 15.

<sup>56</sup>

James R. Parker, *GENerative ART. Algorithms as Artistic Tool*, Alberta 2020, 120–123.

alle digitalen Bildtypen, wobei es Unterschiede in Bildformat und Pixel-Informationen geben kann. Wird das digitale Bild mit einer handelsüblichen Digitalkamera oder einem Scanner produziert, so ist die *Pixel Map* in Form der Auflösung voreingestellt. Die Pixel sind also in einem Koordinatensystem fest verortet und unbeweglich. Auch Bildbearbeitungen verändern die eigentliche Pixelstruktur nicht, sondern wandeln lediglich die Informationen der individuellen Pixel ab.<sup>57</sup> Dies unterscheidet digital gemessene Bilder von den digital generierten Bildern der *GANs*. Denn wie bereits aufgezeigt, ist ein stetiges *Upsampling* entscheidend für ihren Produktionsprozess. Daran anschließend wäre ein naheliegendes Argument für die Unterscheidung, dass die jeweilige Informationsproduktion grundverschieden ist. Während das eine Verfahren durch Halbleitersensoren Lichtimpulse für jedes individuelle Pixel in elektrische Signale umwandelt und so „übersetzt“ und speichert, gewinnt das andere Verfahren Pixel-Informationen durch mathematische Verteilungsberechnungen ohne jede Messung. Diese Annahmen sind jedoch nur teilweise korrekt. Während der in Digitalkameras typische CCD-Sensor die Helligkeitswerte für jedes Pixel der nativen Auflösung misst, gilt dies nicht für die Farbwerte. Hierbei wird typischerweise ein Bayer-Filter verwendet. Dieser Filter misst pro Pixel nur einen der drei RGB-Farbwerte. Um dies auszugleichen, entwickeln die Hersteller Algorithmen, die die Werte zwischen Nachbarn interpolieren.<sup>58</sup> Es lässt sich folgern, dass algorithmische Generierungsprozesse auch ein Teil digitaler Messverfahren sind und keinen grundsätzlichen Unterschied zu digital generierten Bildern bedeuten. Weiter sind beide Prozesse in ihrer Entstehung notwendig singular. *GANs* können deshalb, zumindest gemessen am aktuellen Stand der Technik, niemals absichtlich zwei Gesichter erzeugen, die gleich aussehen. Was bedeutet dies für den referenziellen Charakter generierter Porträts?

Ihr Entstehungsprozess ohne elektronisches Messverfahren impliziert, dass referenzielle Bilder zwar die Grundlage bilden, diese Verbindung aber im durchweg technischen Prozess nicht erhalten bleibt. Denn wie bereits erläutert, ist dieser Prozess kein Kombinationsverfahren: Erzeugt wird ein von Grund auf eigenständiges Bild, das keine Elemente der Trainingsdaten kopiert. Dennoch scheint es – zumindest impliziert dies das Suchverfahren mit PimEyes – eine messbare Übereinstimmung der Bilder zu geben. Der Frage nach potentiellen *Leaks* der Trainingsdaten ging eine Gruppe von Forscher:innen nach:

This paper presents experiments suggesting that identity information in face images can flow from the training corpus

57

Parker, GENerative ART, 138.

58

Tomohiro Hase, Chikako Nakanishi und Tadashi Sakamoto, Software Pixel Interpolation for Digital Still Cameras Suitable for a 32-Bit MCU, in: *IEEE Transactions on Consumer Electronics* 44/4, 1998, 1342–1352, hier 1342–1344.

into synthetic samples without any adversarial actions when building or using the existing model.<sup>59</sup>

Um diese These zu überprüfen, implementierten sie generierte Gesichter und die Bilder der Trainingsdaten in fünf verschiedenen *Face-Matchern*.<sup>60</sup> Ein einheitliches Ergebnis erzielten sie nicht, denn nicht alle der *Matcher* ordneten die realen Gesichter den generierten zu. Da jedoch zwei der Abgleichungstools die echten Gesichter mit den daraus generierten in Verbindung brachten, ist ihre ursprüngliche These der *Identity Leaks* zumindest in Teilen bestätigt. Das Risikopotential veranschaulichten die Forscher:innen in einem Beispiel:

Recently, Facebook shut down thousands of Facebook and Instagram accounts that used ‚AI-generated images‘ as profile photos. If Facebook had chosen to use a matcher behaving as shown [...] [in our test] to hunt down these fake profile photos [...], the real users whose face images appeared in the training data would be 6.6 times more likely to be falsely matched [...].<sup>61</sup>

Diese Ergebnisse zeigen zwar nicht einheitlich, aber zumindest deutlich den real messbaren Nachhall der Trainingsdaten in den generierten Bildern. „So, can we say that these synthetically generated faces truly do not exist?“<sup>62</sup>

Eine undurchlässige Differenzierung der verschiedenen Bildtypen ist demnach nicht haltbar, die fotografische Spur in generierten Bildern nicht verklungen. In ihrem Kern sind sie statistische Annahmen über die durch Trainingsdaten vermittelte Welt, jedoch finden sich in diesen Annahmen Spuren der Ursprungsobjekte, die verortbar sind. Die von *GANs* produzierten Gesichter sind demnach zumindest eine hybride Bildform, die ihr Versprechen gänzlicher Referenzlosigkeit nicht einlösen. Bedenklich ist dies vor allem mit Blick auf die Menschen, die ungefragt zu Trainingsdaten gemacht wurden. Die Rückverfolgbarkeit von Elementen ihrer individuellen Erscheinungen in den generierten Bildern wirft Fragen des Persönlichkeitsrechts auf.<sup>63</sup>

59

Adam Czajka, Patrick Flynn und Patrick Tinskey, This Face Does Not Exist... But It Might Be Yours! Identity Leakage in Generative Models, in: *arXiv*, 1.

60

Dabei handelt es sich um *Deep Learning* Prozesse, die darauf trainiert sind, Gesichter abzugleichen und bei Übereinstimmung zu matchen.

61

Czajka, Flynn und Tinskey, This Face Does Not Exist, 9.

62

Ebd., 1.

63

Die politischen Dimensionen dieser Praxis kommen im abschließenden Kapitel erneut zur Sprache.

Begreifen wir die generierte Bildproduktion als eine erweiterte Form des Fotografischen, und das sollten wir in Anbetracht ihrer Visualität und der bisher geschilderten Verwendung, müssen wir über die damit verbundenen Implikationen nachdenken. Denn wie die generierten Bilder gehandhabt werden, ergibt sich natürlich aus dem Medium, dem sie zu entsprechen versuchen. Die Erwartung wahrheitsliefernder Fotografie hat Pierre Bourdieu in seinem Aufsatz *Die gesellschaftliche Definition der Photographie* treffend beschrieben:

So hat man sich [als partizipierende Gesellschaft] beispielsweise darauf geeinigt, die Photographie als ein Modell der Wahrhaftigkeit und Objektivität zu beschreiben. [...] Es lässt sich nun unschwer zeigen, daß diese gesellschaftliche Vorstellung einer falschen Selbstverständlichkeit aufsitzt. [...] Wenn man die Photographie für die realistische und objektive Aufzeichnung der sichtbaren Welt hält, dann deshalb, weil man ihr (von Anfang an) *gesellschaftliche Gebrauchsweisen* eingeschrieben hat, die als ‚realistisch‘ und ‚objektiv‘ gelten.<sup>64</sup>

Diese von Beginn an eingeschriebene Aufladung, man denke an William Fox Talbots Analogie des *Pencil of Nature*, findet seine Grundannahme in der vermeintlichen Kausalität technischer Bildproduktion.

Zweifellos liegt es ebenso an der gesellschaftlichen Vorstellung vom technischen Gegenstand, der sie hervorbringt, wie an dem gesellschaftlichen Gebrauch, der von ihr gemacht wird, wenn die Photographie als die treue Spiegelung des Wirklichen gilt.<sup>65</sup>

Theoretisiert stand in dieser Diskussion lange das Indexikalische als Wesensbegründung der Fotografie. Rosalind Krauss' Position gilt als wissenschaftlicher Ausgangspunkt dieser Argumentation.

Jede Fotografie ist das Ergebnis eines physikalischen Abdrucks, der durch Lichtreflexion auf eine lichtempfindliche Oberfläche übertragen wird. Die Fotografie ist also eine Form des Ikons, d. h. einer visuellen Ähnlichkeit, die eine indexikalische Beziehung zu ihrem Gegenstand hat.<sup>66</sup>

<sup>64</sup>

Pierre Bourdieu, *Die gesellschaftliche Definition der Photographie*, in: ders., *Eine illegitime Kunst. Die sozialen Gebrauchsweisen der Photographie*, Frankfurt a. M., 85–110, hier 85–86.

<sup>65</sup>

Ebd., 89.

<sup>66</sup>

Rosalind E. Krauss, *Die Originalität der Avantgarde und andere Mythen der Moderne*, Amsterdam 2000, 257.

Sie beruft sich dabei auf den Aufsatz *Die Kunst des Rasonierens* von Charles Sanders Peirce, der das Verhältnis von Gegenstand und fotografischem Abbild aufgrund dieser Kausalbeziehung als indexikalisch wertet.<sup>67</sup> Verschiedentlich gewichtet wurde die so argumentierte Indexikalität zum Kern des fotografischen Wahrheitsversprechen und damit zur Theoretisierung ohnehin gelebter Bildpraxis. Jedoch wurde behauptet, dass das Kriterium der Indexikalität ein Attribut analoger Fototechnik sei, das für die digitale Fotografie nicht mehr in Anspruch genommen werden könne.<sup>68</sup> Doch auch wenn so die indexikalische Theorie und das daran geknüpfte Authentizitätsversprechen zum Distinktionsmerkmal von analoger und digitaler Bildtechnik wird und umgekehrt das digitale Bild „nicht durch die Heilige Dreifaltigkeit von Repräsentation, Index und Punktum erfasst werden“<sup>69</sup> kann, sind die im Analogen verankerten Handlungsweisen ins Digitale übergegangen. W. J. T. Mitchell postuliert, dass

die Frage der Authentizität, des Wahrheitswertes, der Autorität und der Legitimität von Photographien [...] von ihrem Charakter als ‚digitale‘ oder ‚analoge‘ Produktion ganz unabhängig ist. Die Vorstellung, dass der digitale Charakter eines Bildes in einer *notwendigen* Beziehung zur Bedeutung des Bildes steht, zu seinen Wirkungen auf die Sinne, zu seiner Einwirkung auf den Körper oder den Geist des Betrachters, gehört zu den großen Mythen unserer Zeit.<sup>70</sup>

Es sind also, und das knüpft direkt an Pierre Bourdieu an, die Gebrauchsweisen, die das fotografische Bild über sein reines Format hinaus zu eben diesem machen. Selbstverständlich bedeutet dies nicht, dass das digitale dem analogen Naturell gleichzusetzen wäre, sondern verdeutlicht die anhaltende Bedeutungsladung fotografischer Bilder, auch wenn diese ihre physische Existenz überwunden haben.

Dieser so erhaltene fotografische Wahrheitsglaube wird durch das Aufkommen der KI-Bilder ein zweites Mal erschüttert. Dabei ist die Verunsicherung gänzlich kalkuliert und schon im analog-digitalen Übergang angelegt. „Das Ganze des digitalen Bildes ist in Wirklichkeit ein Skeuomorphismus – sein Festhalten an den visu-

67

Charles Sanders Peirce, *Die Kunst des Rasonierens*. Kapitel II (1983), in: ders., *Semiotische Schriften. Band I*, hg. von Christian Kloesel und Helmut Pape, Frankfurt a. M. 1986, 191–201.

68

Beispielsweise hier: Wolfgang Hagen, *Die Entropie der Fotografie*. Skizzen zu einer Genealogie der digital-elektronischen Bildaufzeichnung, in: Herta Wolf (Hg.), *Paradigma der Fotografie*, Frankfurt a. M. 2015, 195–238, hier 233–235.

69

Rubinstein, *Fotografie nach der Philosophie*, 10.

70

W. J. T. Mitchell, *Realismus im digitalen Bild*, in: Hans Belting (Hg.), *Bilderfragen. Die Bildwissenschaften im Aufbruch*, München 2007, 237–256, hier 241.

ellen Konventionen der Fotografie ist schlichtweg eine Designentscheidung [...].“<sup>71</sup> Der Produktionsprozess generierter Bilder spitzt diesen Umstand zu, denn das Bild hat sich weiter vom Referenten entfernt, ohne dies visuell zu kennzeichnen und die visuelle Konvention so zur Selbstreferenz gemacht. Wenn wir diesen Bildtypus als Erweiterung des Fotografischen begreifen, ergibt sich eine doppelte Entmachtung: Zum einen wird die Autorität des Gesichts mitsamt seiner digitalen Kommunikationsstrategien unterwandert. Zum anderen wird hierbei das Wahrheitsversprechen des Fotografischen weiter erodiert. Wie beschrieben haften den generierten Bildern zwar referentielle Artefakte an, doch diese sind im Grunde gegenläufig zu dem, was der Fotografie als Wahrheit zugesprochen wird: die Abbildung von Wirklichkeit. Sicherlich ist dieser Anspruch in der Fototheorie lange schon überwunden, doch in der alltäglichen Praxis ist die visuelle Abbildung der Wirklichkeit noch das zentrale Argument der Fotografie. Keinesfalls meint das einen naiven Glauben ans Bild – die Möglichkeit und Zugänglichkeit von Manipulation sind allgemein bekannt –, sondern vielmehr ein subtiles Versprechen von Authentizität, das dem Medium in seinem digitalen Fluss weiter anhaftet und Beweiskraft zusichert. Von dieser Verantwortung, so könnte man die Technologie als Chance begreifen, kann die Fotografie nun endgültig entbunden werden. Fotografie und Wahrheit sind keine sich bedingende Entitäten und waren es nie. Sie können mit- und ohneeinander, Kausalität darf hierbei aber niemals Voraussetzung sein, sondern muss stets das Ergebnis individueller Aushandlung sein. Blicken wir historisch auf das Fotografische, so ist eindeutig, dass sich die technischen und formalen Parameter des Mediums ständig erweitern. In dieser Kontinuität beanspruchen nun auch generierte Bilder ihren Platz.

## VI. Das fotografische Erbe

Die strenge Formalisierung der generierten Porträts erinnert zunächst an das, was Roland Meyer als *operative Porträts* herausgearbeitet hat. Sie „basieren auf Formaten, die unabhängig vom einzelnen Bild und seinem Gegenstand die Produktion großer Mengen von Bildern regulieren.“<sup>72</sup> Im Falle der generierten Porträts muss jedoch noch einmal zu den Trainingsdaten differenziert werden. Denn bevor das „Bild entsteht, sind dessen Koordinaten bereits festgelegt: Die Abgebildeten begeben sich in einen Raum des Vergleichs, in dem immer schon andere warten.“<sup>73</sup> Das mag rein formal für die generierten Bilder gelten, doch ihre Trainingsdaten sind zum Zwecke der Vergleichbarkeit erst nachträglich forma-

<sup>71</sup>

Rubinstein, *Fotografie nach der Philosophie*, 15.

<sup>72</sup>

Meyer, *Operative Porträts*, 65.

<sup>73</sup>

Ebd.

lisiert und damit operationalisiert worden. In die entsprechende Form gebracht sind sie jedoch genau das: Vergleichsbilder für die rivalisierenden *Neural Networks*. Dass dies nicht im eigentlichen Sinne operativer Porträts ist, liegt auf der Hand. Denn während das operative Porträt eine wie auch immer geartete Identifizierbarkeit gewährleisten soll, verwässert die rein formale Reproduktion des Bildinhalts diese Kernforderung. Für das Porträt in seiner klassischen Form als individuelle Repräsentation muss ein ähnliches Fazit gezogen werden. Wurden die ursprünglichen Bilder auf Flickr noch mit dem Zweck der Vermittlung zwischen „Fremderfassung und Selbstrepräsentation“<sup>74</sup> produziert und gepostet, so können die darauf generierten Bilder dies natürlich nicht mehr leisten. Sie erfüllen eine generische Form, die ausreichend individuell ist, um nicht demaskiert zu werden, alle repräsentativen Attribute außerhalb des Gesichts aber ausspart. Rein formal erfüllen sie zwar die Erfassungsmerkmale von identifikationsdienlicher Fotografie, bleiben darüber hinaus aber ohne Funktionalität und kaum mehr als eine leere Hülle dieses Bildtypus. Der Status des Individuums hat sich in diesen Bildern zu dem einer „nichtexistente[n] Mischperson“<sup>75</sup> geändert, seine Individualität ist nicht mehr als eine folgenlose Implikation. Hans Belting beschreibt diese folgend:

Wo ein Gesicht aus multiplen Gesichtszitaten zusammengesetzt werden kann, entsteht ein virtuelles Gesicht, das sich jeder Referenz auf einen natürlichen Träger, auf ein bestimmtes Gesicht entzieht. [...] Man könnte auch sagen: Es lassen sich Gesichter produzieren, die niemandem gehören, sondern nur noch als Bilder existieren.<sup>76</sup>

Diese vorrausschauende Beschreibung digitaler Porträts erfüllt sich mit den generierten mehr, als das 2013 erahnbar gewesen sein könnte. Das digitale Porträt als Maske beschreibend folgert er:

Im Grunde ist die totale Maske keine Maske mehr, weil da nichts und niemand mehr ist, den sie repräsentiert oder maskiert. Ein digitales Gesicht ist als Bild ein Paradox, weil es die alte Aufgabe der Abbildung von sich weist und mit der Analogie zu einem realen Gesicht seinen historischen Bezug verliert.<sup>77</sup>

<sup>74</sup>

Ebd., 391.

<sup>75</sup>

Richard Weihe, *Die Paradoxie der Maske. Geschichte einer Form*, München 2004, 296.

<sup>76</sup>

Belting, *Faces*, 296.

<sup>77</sup>

Ebd., 298.

Die Frage *wer* da maskiert wird, sofern wir davon ausgehen, dass fotografische Porträts immer eine Maskierung bedeuten,<sup>78</sup> hat sich von der individuellen auf die technische Ebene verschoben. Verschleiert nämlich wird ein gänzlich technischer Produktionsapparat durch die Maske fotografischer Erscheinung. Verstehen wird die Maske als „die Zurschaustellung einer Differenz und damit auch [als] ein Sinnbild für das Erkennen als Differenzieren“<sup>79</sup>, so bedeutet dieses Erkennen eine Verschiebung auf der Bedeutungsebene.

Wird dem Bild die Maske entrissen und sein Ursprung als generiert markiert, schwenkt die Befragung des Individuellen ins Stellvertretende.

Für Künstlerinnen und Künstler stellt sich jedoch immer wieder das Problem, wie Darstellungsaufgaben in Bezug auf das Gesicht ästhetisch zu lösen sind, ohne jene Gespenster wieder auferstehen zu lassen, die mit der Idee der Repräsentation unweigerlich verknüpft sind, wie etwa die Normativität von Aussehen und Geschlecht oder das Konzept der Identität.<sup>80</sup>

Als die Informatiker:innen nun mit der Erfindung Ian Goodfellow's das Feld der Gesichtsproduktion betraten, schalteten sie dieses Problem zumindest theoretisch aus. Für die GANs existiert kein *Gender*, kein Alter und keine *Race*. Diese Konstrukte und Zustände existieren zwar in den Trainingsdaten, sind aber inhaltsleer, da sie zunächst ungelabelt bleiben. Ohne Bedeutungszuschreibung werden sie folglich als statistische Cluster erkannt und reproduziert. Die Lernarchitektur operiert so an den Erscheinungen statistischer Mehrheit, tut dies jedoch ohne binäre Erklärungsmuster. Menschliche Zuschreibungen können sich also gewissermaßen nicht erfüllen, denn die Bilder finden kein kausales Pendant in der Wirklichkeit. Strikter formuliert: Die inhaltsleere Wahrscheinlichkeitsberechnung ist für den technischen Lernprozess frei von biologischer und sozialer Eigenschaft und die einzig anwählbare Kategorie ist das „menschliche Gesicht“. Das kategorisch intentionslose Bild kann durchaus als erfüllt verstanden werden, muss in der Realität allerdings an der Qualität der Trainingsdaten gemessen werden. Denn auch wenn die Technologie eine theoretische Gleichberechtigung verspricht, manifestiert sich in den generierten Bildern soziale Wirklichkeit.

Auffallend häufig blicken Gesichter, die ich als männlich kategorisiere, recht ernst, wohingegen die weiblichen von einem

<sup>78</sup>

Ebd., 120.

<sup>79</sup>

Weihe, *Die Paradoxie der Maske*, 46.

<sup>80</sup>

Ilka Becker, *Verpasste Züge*. Trockel, Prince, Ruff, in: Petra Löffler und Leander Scholz (Hg.), *Das Gesicht ist eine starke Organisation*, Köln 2004, 302–321, hier 304.

Lächeln gezeichnet sind. Die Sichtbarkeit solcher Geschlechterstereotypen lässt darauf schließen, dass sich aufgrund der ebenso geprägten Trainingsdaten von Flickr ein *Algorithmic Bias* gebildet hat.<sup>81</sup> Dieser verdinglicht sich weiter in den generierten Artefakten: Vor allem die weiblich gelesenen Gesichter tragen Schmuck, der in seiner artifiziellen Bildlichkeit wiederum auf den technischen Ursprung der Bilder schließen lässt. Dieser Bias hat den generierten Gesichtern eine soziale Wirklichkeit übergestülpt, welche wiederum ihre Geschichtslosigkeit in Frage stellt. Es muss hier festgestellt werden, dass der Output nur so gut sein kann, wie das Trainingsmaterial. Denn wenn dieses schon korrumpiert ist, setzt sich die Verschiebung in den *Hidden-Layers* der Lernarchitektur fest und reproduziert sich schlussendlich in den Bildern. Den größten Effekt auf die Qualität des Outputs haben Über- und Unterrepräsentationen definierbarer Gruppen.<sup>82</sup>

Eindrücklich belegt dies die Studie *AI-synthesized faces are indistinguishable from real faces and more trustworthy*.<sup>83</sup> Die wenig überraschende Erkenntnis dreier Experimente ist, dass die Bilder für die Partizipant:innen nicht zuverlässig von fotografischen Porträts unterscheidbar waren, sofern diese die gleichen formalen Parameter erfüllen.<sup>84</sup> Durch die Diversifizierung der Bilder in Alter, *Race* und *Gender* zeigte sich zudem, dass *weiße* Männer am unzuverlässigsten als generiert erkannt wurden, die Erklärung hierfür liegt schlicht in ihrer Überrepräsentation in den Trainingsdaten.<sup>85</sup> Dort wo mehr Daten zum Training vorhanden sind, präzisiert sich das *GAN* in seinem Output. Umgekehrt bedeutet hier die Unterrepräsentation eine stärkere Varianz in der Bildqualität. So reproduzieren sich Effekte von *Gender* und *Race*, insofern als dass Unter- und Überrepräsentation an diesen Parametern festzustellen sind, in den generierten Bildern.<sup>86</sup> Neben diesen durchaus erwartbaren Ergebnissen präsentiert die Studie ein verblüffendes Ergebnis: Die generierten Gesichter wurden als in höherem Maße *trustworthy*

81

Zum *Algorithmic Bias* siehe auch: Meyer, Gesichtserkennung, 27–28.

82

Sophie J. Nightingale und Hany Farid, AI-synthesized faces are indistinguishable from real faces and more trustworthy, in: *PNAS* 119/8, 2022, 3.

83

Ebd.

84

Bei den Vergleichsbildern handelte es sich um den Trainingsdatensatz von *StyleGAN2*. Identisch waren deshalb: Auflösung, quadratischer Bildausschnitt, Position des Gesichts und insbesondere der Augen.

85

Nightingale und Farid, AI-synthesized faces, 1.

86

Ebd.

gelesen als die Gesichter auf den fotografischen Porträts.<sup>87</sup> Warum die Studienteilnehmer:innen in der Mehrzahl generierte Porträts als vertrauenswürdiger markierten, lässt sich im Detail bisher nicht erklären. Da mehr als subtile Implikationen Auslöser des überwiegend einheitlichen Verhaltens nicht sein können, ist ein Spekulieren darüber an dieser Stelle auch nur bedingt zielführend. Sinnvoller ist dagegen, auf die gemessene Ungleichheit in der Repräsentation zurückzukommen.

Der Abgleich qualitativer Übereinstimmung generierter Porträts mit dem menschlichen Gesicht ist stets Qualitätsmarker der Trainingsdaten. Umgekehrt kann der Diskriminator eine Norm definieren, die alles ablehnt, was unter bestimmten Gesichtspunkten von der Masse abweicht. Auch hier ist diese Varianz maßgeblich durch den Umfang und die Diversität der Trainingsdaten definiert. Die Gefahr ist allerdings, dass dies nur sichtbar wird, wenn die Gesamtheit der Trainingsdaten bekannt ist und abgeglichen werden kann. Ist dies nicht gewährleistet, können Marker und Identitäten schlichtweg unsichtbar werden, weil die Cluster für die Lernarchitektur zu insignifikant sind. Ähnlich subtil setzen sich diskriminierende Vorstellungen fort, die technisch in den Trainingsdaten eingeschrieben sind. Es ist mittlerweile hinlänglich belegt, dass die technische Geschichte der Fotografie nicht zuletzt eine der *weißen* Normiertheit ist:

Die technische Prädisposition der Filmempfindlichkeit auf den Reflexionsgrad *weißer* Haut konnte in Einzelportraits von Schwarzen Menschen durch kompensatorische Maßnahmen wie Zusatzbeleuchtung bei der Aufnahme und anschließende labortechnische Feineinstellungen nach dem Prinzip von Versuch und Irrtum ausgeglichen werden.<sup>88</sup>

Die fotografische Darstellungsqualität Schwarzer Menschen war faktisch schlechter als die *weißer* Menschen, die normierter Ausgangspunkt der technischen Kalibrierung waren. Durch Bewusstsein um diesen Missstand und technischer Entwicklungen im Hinblick auf Film- und Sensorempfindlichkeit im Digitalen hat sich diese Normierung zwar deutlich abgeschwächt, jedoch nicht gänzlich aufgelöst. Durchaus finden sich die Effekte *weißer* Privilegierung in die schwer einsehbaren Bildproduktionsprozesse digitaler Kameras und den Algorithmen der Bildverarbeitung eingeschrie-

<sup>87</sup>

Race spielte hierbei keine Rolle, weiblich gelesene Personen wurden jedoch als vertrauenswürdiger gewertet.

<sup>88</sup>

Jakob Schnetz, Beharrliche Apparate. Eine rassismuskritische Untersuchung weißer Normen in fotografischer Technologie und ihrer algorithmischen Nutzung, in: Rebecca Ramershoven und Simon Dickel (Hg.), *Alles Uns. Differenz, Identität, Repräsentation*, Münster 2022, 64–93, hier 69.

ben.<sup>89</sup> Notwendigerweise resultiert dies in ungleicher Repräsentation: „Neben der ikonografischen ist ebenso die technische Repräsentation Schwarzer Haut stets ein identitätspolitisches Moment, das sich gegenüber impliziten wie expliziten *weißen* Normen zu behaupten hat.“<sup>90</sup> Diese in der technischen Blackbox digitaler Bildgebungstechnik verborgenen Normierungen setzen sich ungebrochen in die Lernarchitektur der *GANs* fort. Gewissermaßen erfährt die Normierung hier eine zweite Normalisierung, denn gemeinhin wird KI-Anwendungen ein hohes Maß an objektiver Datenproduktion zugesprochen. Dabei ist längst belegt, wenn auch weniger bekannt, dass Technologie alles andere als neutral ist:

Die Schattenseite der ‚Innovation‘ und der verbesserten ‚Effizienz‘ automatisierter Technologien liegt darin, wie sie marginalisierte Gruppen – rassifizierte Menschen, Migrant:innen ohne Papiere, queere Communities und Menschen mit Behinderungen – kategorisieren und mit ihnen experimentieren.<sup>91</sup>

Basieren Entscheidungsprozesses auf den Empfehlungen künstlicher Intelligenz, so laufen diese Gefahr, in diskriminatorischen Praktiken zu resultieren. Denn wenn die Trainingsdaten schon Diskriminierungsmuster aufweisen, erfahren diese durch die maschinelle Auswertung nur Objektivierung und sorgen so für den Erhalt alter Macht- und Herrschaftssysteme.<sup>92</sup> Im Falle der generierten Porträts müssen wir anerkennen, dass das Gesicht politisch ist: Zum einen, weil schon die Trainingsdaten Ungleichheit in sich tragen, zum anderen aufgrund ihrer technologisierten Reproduktion.

Diese Erkenntnis erfährt ihre Gewichtung vor allem dann, wenn die jüngsten Fortschritte der *GAN*-Technologie miteinbezogen werden. Die frühe *StyleGAN*-Architektur, sowie die Bilder auf der Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) operierten noch ohne Kategorien. Parameter wie *Gender*, *Alter* oder *Race* konnten nicht zur maschinellen Kategorisierung der produzierten Gesichter genutzt werden, denn zumindest auf der Nutzeroberfläche existierten diese Label nicht. Da nun die Kategorie „menschliches Gesicht“ in einer Porträtsammlung trivial ist und die engste faktische Kategorie „Porträts von Flickr“ nicht mehraussagend ist, sind

89

Ebd., 81.

90

Ebd.

91

Sarah Chander, Datenrassismus. Eine neue Ära, in: [Netzpolitik.org](http://Netzpolitik.org), 29.02.2020 (30.09.2023).

92

Folgende Bücher möchte ich hier zur Vertiefung empfehlen: Ruha Benjamin, *Race after Technology. Abolitionist Tools for the New Jim Code*, Cambridge 2019. Wendy Hui Kyong Chun, *Discriminating Data. Correlation, Neighborhoods, and the New Politics of Recognition*, Cambridge, MA 2021. Safiya Umoja Noble, *Algorithms of Oppression. How Search Engines Reinforce Racism*, New York 2018.

die Bilder intrinsisch als Typenbilder nicht zu gebrauchen. Dieses dennoch unentschlossene Stadium zwischen Individuum und Typ hat sich mittlerweile zugunsten des Typs verschoben. Als Typenbilder sind in der Fotografie solche zu verstehen, die stellvertretend für eine reale oder konstruierte Gruppe stehen und visuell eine Aussage über diese implizieren. Ein Musterbeispiel sind die von Francis Galton entwickelten Kompositfotografien. Universeller, doch in gleicher Fahrtrichtung ist das Werk *Antlitz der Zeit* (1929) von August Sander zu werten. Der Porträtfotograf versuchte durch Typenbilder einen visuellen Querschnitt der Gesellschaft zu erschaffen. Denn „wo er seine ‚Typen‘ [...] an ihrem Arbeitsplatz zeigt, präsentiert er uns [...] Figuren, die mitsamt ihrer Attribute als exemplarische Vertreter eines Berufsstandes posieren.“<sup>93</sup> Auch kennen wir diese Bildargumentation aus dem kolonialistischen Kontext mit visuell rassistischen Implikationen sowie physiognomischen Bildbefragungen, wie sie nicht zuletzt im Dritten Reich Verwendung fanden. Es sind im Einzelbild angelegte Inszenierungen, die sich als Kollektiv vervollständigen und so universelle Aussagen provozieren. Anzunehmen, dass derlei fotografische Befragungen der Vergangenheit angehören würden, ist insbesondere im Kontext der KIs falsch, denn gerade im „Zeitalter von Big Data und Deep Learning erfahren längst überwunden geglaubte physiognomische Traditionen eine Renaissance“.<sup>94</sup> Die hier diskutierten generierten Porträts erweckten den Anschein, derartige Kategorisierungen zu unterwandern, weil das Lernsystem zunächst ohne Labels operierte. Jedoch wird es wenig überraschen, dass das kategorische Labeln mittlerweile zur gängigen Praxis generierter Bilder gehört. Auf der Website [www.this-person-does-not-exist.com](http://www.this-person-does-not-exist.com), die eindeutig an die bisher analysierte Website [www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com) angelehnt ist, können Porträts nach den Kategorien Alter, *Gender* (binär) und *Race* ausgewählt werden.<sup>95</sup> So erhebt sich das digitale Kollektivporträt zum Typenbild und die referenzlosen Gesichter werden zu Stellvertreter:innen der angewählten Kategorien. Dem alten fotografischen Typenfetisch folgend haben die Entwickler:innen das technologische Potential grenzenloser Gesichter beendet. Bedenken wir die zuvor erörterten strukturellen Probleme der Bildgebungstechnik, sollten wir vorsichtig sein, mit ebendiesen Kategorien zu arbeiten. Immerhin setzen sich subtil die realen Machtverhältnisse in den generierten Porträts fort. Spätestens hier ist klar, dass diese

93

Meyer, Operative Porträts, 258.

94

Ders., Gesichtserkennung, 34–35.

95

Eine vergleichbare Website ist [Generated.Photos](http://Generated.Photos). Bei beiden Websites muss man für Bildproduktion ohne Wasserzeichen bezahlen.

Bilder keinesfalls so frei von Bezügen zur existierenden Welt sind, wie ich ihnen eingangs unterstellt habe.<sup>96</sup>

## ANHANG

- <https://freehookup.dating/online-dating/us/texas/fort-worth/divorced/single.php>
- <https://www.sexlugar.es/gays/linares.html>
- [https://flirthookup.dating/catholic-singles-fl-vero\\_beach.html](https://flirthookup.dating/catholic-singles-fl-vero_beach.html)
- <https://pistolasdemasaje.top/ekupuz/>
- <https://melhores-aplicativos.com/305/reunioes-ringcentral>
- <https://www.realforexreviews.com/author/brianbgrounds>
- <https://freehookup.dating/online-dating/us/north-carolina/durham/jewish/single.php>
- <https://besthookup.dating/member338051.htm>
- <https://www.purobrillo.es/>
- <https://besthookup.dating/member114524.htm>
- <https://tuinplanten.wiki/2435/ongedierte-dat-op-taxusbomen-kan-voorkomen>
- <https://datehookup.dating/user-180620909.htm>
- <https://datehookup.dating/singles-la-farmerville.htm>
- <https://datehookup.dating/Single-Men-PA-Jeannette.htm>
- <https://melhores-aplicativos.com/9954/ring4>
- <http://www.aytoloja.org/jforum/user/profile/140311.page;jsessionid=1B30D1DFCAFACD95B8E05096C5E20267>
- <https://thaiwomen.dating/Nillawan>
- <http://bteeosw.online/page/londyn-letiste-gatwick/>

**Paul Werling** machte seinen Bachelor in Kunstgeschichte, Filmwissenschaft und Soziologie an der Friedrich-Schiller-Universität Jena. 2023 schloss er den Master im Fach *Photography Studies and Research* an der Folkwang Universität der Künste in Essen ab. Thematisch beschäftigt er sich mit digitalen Aspekten des Fotografischen sowie Fotografie im Kontext von Unrechtsregimen.

96

Ganz herzlich möchte ich mich bei den anonymen Peer-Reviewer:innen für ihre wertvollen Hinweise und ihre Kritik bedanken. Mein Dank gilt weiter den Herausgeber:innen und der Redaktion von *2t: Inquiries into Art, History, and the Visual* für die Veröffentlichung und Aufbereitung meines Textes. Zuletzt bedanke ich mich bei Steffen Siegel, Roland Meyer und Markus Rautzenberg für ihre Anmerkungen und Hilfe beim Verfassen dieses Artikels.