

# Open and Closed AI. Eine Kunstkritik künstlich generierter Bilder

**Bell, Peter**

peter.bell@uni-marburg.de

Philipps-Universität Marburg, Deutschland

## Einleitung

Unter dem Hashtag #aiart erscheinen täglich tausende künstlich erzeugte Bilder auf sozialen Medien und digitalen Marktplätzen z.B. als NFTs, flankiert von Kommentaren und Artikeln, welche die Qualität und Kreativität dieser Bildproduktionen diskutieren.

Der Vortrag reflektiert diese Entwicklungen, indem er die Bilder und ihre Genese kritisch hinterfragt, wertet und die Prozesse analysiert. Dass Digital Humanities parallel zum wissenschaftlichen Rezensionswesen auch eine Kritik von Software entwickeln sollte, zeigt sich z.B. in der Gründung des CKIT Rezensionsjournal das aus dem AK digitale Kunstgeschichte und nfdi4culture entstanden ist. Im Zusammenhang generativer künstlicher Intelligenz weitet sich diese Softwarekritik in Form einer Werkkritik aus. Aufgrund der raschen Entwicklung innerhalb der Computer Vision und der generativen Modelle gibt es bislang wenige Reflexionen über deren Schöpfungshöhe (Gary et al. 2022) jenseits der informatischen Überbietungslogiken und der stetigen Diskussion über bias von AI.

## OpenAI und closed access. Technikkritik

Digitale Kunstkritik setzt bereits ein, wo die Frage gestellt wird, ob dieses popkulturelle Phänomen Gegenstand der Kunstgeschichte oder der Medienwissenschaften ist oder – wie öffentlich viel diskutiert – es sich überhaupt um Kunst handeln kann. Die Frage ist leicht zu beantworten, denn in dem Moment, wo über Kunst oder nicht Kunst entschieden werden muss, bedarf es unvermeidlich der Methoden der Kunstgeschichte und Kunstkritik. #Aiart könnte auf einer rein ästhetisch-visuellen Ebene mit konventionellen Methoden der Kunswissenschaft bewertet werden. Dadurch, dass die Bilder nicht von Menschen sondern in erster Linie von einem mathematischen Modell erzeugt werden, bietet sich an eine operative und technikzentrierte Medienkritik in die Kunstkritik zu integrieren. Die hier durchzuführende kritische Reflexion über AI Art findet somit in der digitalen Kunstgeschichte als Teil der Digital Humanities statt. Dies geschieht in Form Medienarchäologie,

in der das maschinelle Lernen und die verschiedenen Modelle mit- und gegeneinander arbeitender neuronaler Netze untersucht werden. Dabei zeigt sich eine Vielzahl an Architekturen von GANs und Transformer-Netzwerken<sup>1</sup> deren Ergebnisse teils zusätzlich noch durch verschiedene Trainingsdatenbanken<sup>2</sup> variiert werden können. Die 2014 mit generative adversarial networks (GAN) begonnene Entwicklung hat sich in den letzten drei Jahren deutlich beschleunigt und zu immer realistischeren Bildern geführt. Während einige Modelle selbstständig Bilder erzeugen, ist die Eingabe via Bildbeschreibung quasi als Kommando (prompt) zum Standard der künstlichen Bildproduktion geworden. Diese Texteingabe ist möglich, da in DALL-E ein Sprachmodell (GPT-3) mit Algorithmen zur Bilderzeugung kombiniert ist. Auch durch dieses Zusammenkommen von Bildverarbeitung/Visualisierung und Computerlinguistik besteht hier für die Digital Humanities ein großes Potential an Forschung, Anwendung und Kritik. Daneben verbinden sich durch die Bildsynthese die Forschungsfelder Computer Vision und Computer Grafik noch enger, so dass ein transdisziplinärer bildwissenschaftlicher Gegenstand vorliegt. Eine erste Phase von ‚Kunstkritik‘ ist den Architekturen schon eingeschrieben, da ständig entschieden wird, welche Bilder den Trainingsdaten ähneln oder durch andere Kriterien als angemessen erscheinen.

DALL-E der Stiftung OpenAI lässt sich durch sieben Gründe nicht als ‚open‘ bezeichnen, da der Code ist bislang nicht publiziert worden und die Daten anhand derer das maschinelle Lernen durchgeführt wurde (oder werden) nicht bekannt sind. Bekannt ist lediglich, dass es sich um im Internet gescrapete Bild- und Textkombinationen (z.B. Bild und Bildunterschriften) handelt. Zudem gab es zunächst keinen freien Zugang zur Nutzung, sondern nur eine Registrierung über eine Warteliste und keine Transparenz über die Zugangsvergabe. Seit 20.07.2022 ist DALL-E monetarisiert, indem nach einer gewissen Anzahl unentgeltlicher Bilder, dann für weitere Bilderzeugungen bezahlt werden muss. Ähnlich wie bei Midjourney erscheinen diese Kosten allerdings relativ moderat. Die generierten Bilder gehören alle openai. Selbst wenn ein eigener Bildupload die Grundlage bildete oder das Kommando eine Innovation/Schöpfungshöhe besitzt. Allerdings räumt openai seit Juli 2022 eine Nutzung der Bilder auch zu kommerziellen Zwecken ein. Gleichzeitig wehren sich Künstler\*innen international gegen die unfreiwillige Verwendung ihrer Bilder als Trainingsdaten, da ihr Stil so sehr einfach reproduzierbar wird.

Die Offenheit von DALL-E wird zuletzt durch eine teils rigide Content Policy eingeschränkt, wodurch Bilder mit sensiblen Inhalten oder Deepfakes verhindert werden sollen. Andere Modelle wie Stablediffusion sind offen im Hinblick auf open source, open access und freier Texteingabe, was wiederum Kritik aufgrund der Missbrauchsmöglichkeiten erzeugt.

Eine kritische Reflexion darüber, ob OpenAI mit dieser Herangehensweise nicht ihrer eigenen Stiftungsscharta<sup>3</sup> widerspricht, kann nur angerissen werden. Die Geschlossenheit der kommerziellen Systeme wie DALL-E, IMAGEGEN und PARTI ist allerdings nur ein Aspekt, der eine wissenschaftliche Erforschung und Kritik erschwert. Der Aufbau konkurrenzfähiger Modelle an Universitäten scheitert oft am Aufbau der Trainingsdatensätze sowie

der Infrastruktur respektive Rechenleistung für das maschinelle Lernen. Die akademische Forschung hat somit Schwierigkeiten selbst unabhängige Architekturen zu erstellen und kann gleichzeitig mangels Verständnisses der führenden Modelle keine fundierte Kritik zu deren Funktionsweise artikulieren. Modelle wie Stablediffusion (Rombach 2022) zeigen hingegen, dass durchaus konkurrenzfähige Architekturen im akademischen Kontext entstehen klönnen. Insgesamt besteht allerdings die Herausforderung die Verfahren der oft als black box beschriebenen Netze zu interpretieren. Dies ist u.a. möglich durch den Vergleich der unterschiedlichen Modelle und Trainingsdatensätzen oder Ergebnissen aus unterschiedlichen Phasen der Bildgenerierung, sowie durch Prompt Engineering als experimentelle Methode. Bei letzterer kombiniert und ergänzt man in Versuchsreihen die Satzteile einer Befehlszeile in unterschiedlichen Reihenfolgen. Dabei können auch Marker gesetzt werden, z.B. indem im Kommando dazu aufgefordert wird, Personen einzufärben oder auf andere Weise Objekte zu markieren, um zu sehen, ob das Modell nicht nur Bilder des Konzepts kopiert, sondern ein tieferes Verständnis des Konzepts hat. Auch kann untersucht werden, wie ein Bild durch Zugabe oder Umformulierung von Text realistischer wird. An dieser Stelle zeigt sich auch wie subjektiv die Kriterien der Bewertung sind. Welche Grundlage hat die Wahrnehmung ein Bild realistischer oder die Umsetzung (man könnte auch sagen Inszenierung) eines prompts gelungener zu finden?

## Kunstkritik

Neben der Medienarchäologie zu neuronalen Netzen und dem maschinellen Lernen muss hier also eine zweite Form der Auseinandersetzung gefunden werden, die sich methodisch an klassische Kategorien und Werte der Kunstkritik anlehnt. Diese Werte sind nicht essentialistisch zu verstehen, sondern relativistisch. So kann ein visueller Turing-Test durchgeführt werden, indem gefragt wird, ob das Bild für computergeneriert oder eine Fotografie bzw. von Menschen digital oder konventionell erstellte Grafik gehalten wird. Hier zeigt sich, dass in DALL-E 2 viele Bilder erzeugbar sind (Photograph of a house in the Bauhaus style; Abb. 1), die in diesem Turing-Test bestehen, während andere unbefriedigende Ergebnisse liefern (17th century landscape painting). Das Problem ist in vielen Fällen nicht die Imitation eines Stils, sondern die Form der dargestellten Dinge also die Detailschilderungen (im Fall der niederländischen Landschaftsmalerei sind es monströse Kühe). Die Bildgeneratoren insbesondere DALL-E 2 sind intelligent im Sinne von Mimikry. Sie können die Oberflächen von Dingen, Stilen, Texturen und Beleuchtungen überzeugend wiedergeben ohne ein tieferes Verständnis der Gegenstände zu haben. Die für die Kunstgeschichte und Kunstkritik so wichtigen Pole, Kunst und Natur, haben auch bei der künstlichen Bildgenerierung eine Bedeutung. Das antrainierte „Weltwissen“ führt dazu, dass die Modelle auch in der Natur vorhandene Proportionen internalisieren – also visuelle Prinzipien der Kompositionslässtigkeit befolgen. Entsprechend fällt es beispielsweise dem Algorithmus schwer Elefanten zu erzeugen, die kleiner als Schildkröten sind. Durch erscheint auch der Namensgeber Salvatore Dali

für DALL-E eher unpassend, weil dessen radikal unnatürlichen Schöpfungen innerhalb seiner surrealen Kunst mit DALL-E kaum zu reproduzieren und noch weniger von diesem Modell zu erfinden wären.

Auf diese Weise sollen Kriterien bzw. die so genannten Rubriken der Kunstkritik (Vogt 2010) wie Bilderfindung, Komposition, Ausdruck, Stil und Dekorum, Naturnähe (im Sinne von real-world-data) und Kunstrezeption untersucht werden. Auf ikonologischer Ebene führt die Bildkritik direkt in gesellschaftliche Zusammenhänge. Denn auch bei den neuen Generationen von Bildgeneratoren zeigen sich der Bias, der in Bezug auf das maschinelle Lernen bereits viel besprochen wurde (z.B. Stereotype Frauen- und Männerberufe sowie nicht-weiße Kriminelle werden wie selbstverständlich erzeugt), während die vollständige Unwissenheit bzgl. einiger kanonischer Werke der Kunst- und Kulturgeschichte auch die mangelnde Historizität der Trainingsdaten aufzeigt.

Anhand der exemplarischen Beurteilung der synthetischen Bilder soll es weniger um Urteile der einzelnen Bilder gehen, als um die Aufstellung von Kriterien und Erkenntnisse über die Potentiale von aiart insgesamt. Dabei zeichnet sich ab, dass wir es mit einer Form von Mimikry zu tun haben, durch die Oberflächen und Stile immer besser imitiert werden können, während komplexe Kompositionen und Konzepte scheitern. Die Forschungen zu aiart entstanden im Rahmen des SPP „Das digitale Bild“ im Projekt „Bildsynthese als Methode des kunsthistorischen Erkenntnisgewinns“ (2019–2022). In Ausblick und Diskussion soll auch darauf eingegangen werden, welche weiteren Anwendungsfelder sich in der Kunstgeschichte für die künstlich generierten Bilder aufzeigen lassen.



Abb. 1: Mit Dall-E 2 erzeugte Bauhaus Architekturen.

## Fußnoten

1. StyleGAN, VQGAN+CLIP, GAUGAN, craiyon, DALL-E 1 und 2, Imagegen, Midjourney, Stablediffusion.
2. z.B. ImageNet, OpenImage, CelebA.
3. OpenAI Charter: <https://openai.com/charter/>

## Bibliographie

**Chen, Weiwen, Mohammad Shidujaman, und Xuelin Tang** (2020). AiArt: Towards Artificial Intelligence Art. In The 12th International Conference on Advances in Multimedia.

**Esser, Patrick, Robin Rombach, und Björn Ommer** (2021). *Taming Transformers for High-Resolution Image Synthesis*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12873-12883.

**Kohle, Hubertus, und Stefan Germer (1991)**. „Sponsaneit t und Rekonstruktion. Zur Rolle, Organisationsform und Leistung der Kunstkritik im Spannungsfeld von Kunsttheorie und Kunstgeschichte“. Buchbeitrag. Wiesbaden. <https://archiv.ub.uni-heidelberg.de/artdok/109/>.

**Marcus, Gary, Ernest Davis, und Scott Aaronson** (2022). *A very preliminary analysis of DALL-E 2*. arXiv, 2. Mai 2022. <https://doi.org/10.48550/arXiv.2204.13807>.

Offert, Fabian, und Peter Bell (2022). *Generative Digital Humanities*. In CHR, pp. 202-212.

**Oppenlaender, Jonas** (2022). *Prompt Engineering for Text-Based Generative Art*. preprint arXiv:2204.13988.

**Rombach, Robin, et al.** (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 10674-85, <https://doi.org/10.1109/CVPR52688.2022.01042>.

**Saharia, Chitwan, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, et al.** (2022). *Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding*. arXiv. <https://doi.org/10.48550/arXiv.2205.11487>.

**Sparkes, Matthew** (2022). *AI Art Tool Covertly Alters Requests*. New Scientist 255, Nr. 3397 (30. Juli 2022): 10.

**Vogt, Margrit** (2010). *Von Kunstworten und -werten: Die Entstehung der deutschen Kunstkritik in Periodika der Aufkl rung*. De Gruyter, Berlin.